PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : A61K 38/00, 39/00, 39/395, C07H 21/04, C07K 1/00, 16/00, C12N 5/00, 7/00, 15/09, C12P 21/08, G01N 33/53

A1

(11) International Publication Number: WO 00/61165

(43) International Publication Date: 19 October 2000 (19.10.00)

(21) International Application Number: PCT/US00/09866

(22) International Filing Date: 13 April 2000 (13.04.00)

(30) Priority Data:
60/129,073 13 April 1999 (13.04.99) US

(71) Applicants *(for all designated States except US)*: SMITHKLINE BEECHAM CORPORATION [US/US]; One Franklin Plaza, Philadelphia, PA 19103 (US). SMITHKLINE BEECHAM PLC [GB/GB]; New Horizons Court, Great West Road, Brentford, Middlesex TW8 9EP (GB).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: LUPAS, Andrei, Nicolae [DE/US]; 487 Mockingbird Lane, Collegeville, PA 19426 (US).

(74) Agents: GIMMI, Edward, R. et al.; SmithKline Beecham Corporation, Corporate Intellectual Property, UW2220, 709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406–0939 (US).

(81) Designated States: JP, US, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: CONSERVED ADHESIN MOTIF AND METHODS OF USE THEREOF

(57) Abstract

Isolated polypeptides which are conserved in eubacterial extracellular domains are identified in five pathogens of the beta and gamma branches of proteobacteria. These polypeptides, alone or as fusion proteins with a second protein, are useful in the generation of antibodies or other antagonists. The peptides, fusion proteins, and antibodies are useful as vaccine components or therapeutic agents against bacterial infection or as diagnostic reagents. These polypeptides are also useful in screening methods for other agonists and antagonists which may be used in diagnosis, therapy, and as vaccines.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

CONSERVED ADHESIN MOTIF AND METHODS OF USE THEREOF

Field of the Invention

This invention relates to newly identified polynucleotides, polypeptides encoded by such polynucleotides, the use of such polynucleotides and polypeptides, as well as the production of such polynucleotides and polypeptides. The polypeptides of the present invention have been putatively identified as conserved binding domains of bacterial adhesins.

Background of the Invention

Adhesins are multifunctional, proteinaceous structures formed on the surface of pathogenic bacteria. They can mediate attachment or can assist the invading bacteria in avoiding the immune responses of the infected host, which are intended to protect the host against the infection. Adhesins can interaction with host cell receptors, in adherence to extracellular matrix proteins (ECM), in activation or inactivation of host proteases (e.g. activation of plasminogen or inhibition of complement).

The virulence plasmid (pYV)-encoded non-fimbrial surface protein YadA (Yersinia adhesin, formerly known as Yop1 or P1) of the enteropathogenic *Yersinia enterocolitica* and *Y. pseudotuberculosis* is an important virulence determinant of the bacterial enteropathogenic *Yersinia* species [A. Roggenkamp *et al*, Infec. Immun., 64(7):2506 (July 1996)]. The adhesin has several functions. It is crucial for pathogenicity; it mediates adherence to epithelial cells, professional phagocytes and ECMs; it binds to the complement inhibitor factor H and appears to protect the bacterium against complement and defensin lysis [Cornelis and Wolf-Watz, Mol. Microbiol., 23:861-867 (1997); Heesemann and Grüter, FEBS Microbiol. Lett., 40:3-41 (1987); Roggenkamp *et al.*, Mol. Microbiol., 16:1207-1219 (1995) Roggenkamp *et al.*, (1996) cited above; Visser *et al.*, Infec. Immun., 64:1653-1658 (1996)]. YadA is also involved in auto-agglutination, a phenomenon occurring after growth in tissue culture medium at 37EC [Skurnik *et al.*, J. Bacteriol., 158:1033-1036 (1984)].

Despite this knowledge of its structural and functional characteristics. there is nothing known about the sequence of this adhesin that would permits its use diagnostically, therapeutically or prophylactically. If compositions could be developed to interfere with the functions of the adhesin, however, such compositions and methods of use thereof would prove useful in the therapeutic and prophylactic treatment of the bacterial infections mediated by these pathogen. Thus, there exists a need in the art for proteins, antagonists and agonists of these bacterial adhesins, as well as compositions and methods for their use in the vaccine and diagnostic fields.

Summary of the Invention

In one aspect, the present invention provides isolated polypeptides of about 20 amino acids in length, which are conserved in proteobacterial extracellular proteins and which bind to a protein or proteinaceous ligand expressed by a mammalian cell. In a particularly preferred embodiment of this aspect of the invention, the polypeptides of this invention include the sequences of the invention set forth herein, which are found in *Neisseria*, *Actinobacillus*, *Haemophilus*, *Moraxella* and *Yersinia* pathogens. Biologically active and diagnostically or therapeutically useful fragments, variants, analogs and derivatives of these sequences are provided, as well as variants and derivatives of the fragments, and analogs of the foregoing. These polypeptides, which are free from association with other contaminating or proteinaceous materials with which they are found in nature, may be produced synthetically or by recombinant means.

In another aspect of the present invention, there are provided non-naturally occurring synthetic, isolated and/or recombinant polypeptides, fragments, consensus fragments and/or sequences having conservative amino acid substitutions of the conserved proteobacterial sequences of the present invention. These polypeptides may bind proteobacterial adhesin ligands, or may also modulate, quantitatively or qualitatively, adhesin ligand binding.

In another aspect, the present invention provides synthetic, isolated or recombinant polypeptides which are designed to inhibit or mimic various conserved proteobacterial adhesin sequences or fragments thereof.

In another aspect, the invention provides a fusion protein which comprises

at least one of the polypeptides of this invention fused in frame to a second protein.

In still another aspect, the invention provides an isolated or synthetic polynucleotide sequence free from association with other materials with which it is found in nature, which encodes a polypeptide or fusion protein described herein. including nucleic acid probes comprising nucleic acid molecules of sufficient length to specifically hybridize to nucleic acid sequences of the present invention.

In yet a further aspect. the invention provides a nucleic acid molecule, e.g., a vector or plasmid or recombinant virus, comprising a polynucleotide sequence encoding a polynucleotide sequence or fusion protein of this invention under the control of regulatory sequences which direct the expression of the polypeptide or fusion protein in a host cell.

In a further aspect. the invention provides a host cell comprising the nucleic acid molecule described above.

In another aspect, the invention provides a composition which inhibits or retards the binding of a proteobacterial adhesin to its ligand or to a cell expressing its ligand.

In another aspect of the invention, there are provided antibodies which bind to the conserved polypeptides, including humanized antibodies, anti-antibodies, monoclonal and polyclonal antibodies, among others. In still another aspect, the invention provides an anti-idiotype of the antibody described above.

In yet a further aspect, the invention provides an immunogenic composition useful as a vaccine to prevent infection by a proteobacterial species comprising in a pharmaceutically acceptable carrier a polypeptide or fusion protein of this invention. In one embodiment, this composition contains a conserved proteobacterial polypeptide or fusion protein of this invention, or an immunogenic fragment thereof. In another embodiment, the composition contains an amino acid sequence at least 70% identical to the aforementioned sequences as determined by a sequence comparison algorithm, which sequence binds the adhesin ligand. In another embodiment, the composition contains a small molecule which binds the adhesin ligand. In yet a further embodiment, the composition contains an antibody which binds the adhesin ligand, or an anti-idiotype antibody of that antibody. These compositions may also contain one or more adjuvants or

carriers.

In a further aspect, the invention provides a process for producing the aforementioned polypeptides, polypeptide fragments, variants and derivatives, fragments of the variants and derivatives, and analogs of the foregoing. In a preferred embodiment, the invention provides methods for producing the aforementioned polypeptides by recombinant techniques comprising culturing recombinant prokaryotic and/or eukaryotic host cells, containing (i.e., having expressibly incorporated therein) a nucleic acid sequence encoding a polypeptide of the present invention under conditions for expression of the polypeptide in the host and then recovering or isolating the expressed polypeptide from the cell or cell lysate. In another embodiment, the polypeptides of this invention are produced by conventional synthesis methods.

In still another aspect, the invention provides a method for vaccinating a mammalian subject against infection by a proteobacteria which includes administering to the subject a prophylactically effective amount of the immunogenic composition described above.

In another aspect, the invention provides a method of making an immunogenic composition for use as a vaccine component against proteobacterial infection comprising fusing a polypeptide of this invention to a second protein capable of resisting degradation *in vivo*, wherein said polypeptide elicits antibodies *in vivo* which interfere with the binding of proteobacterial adhesin molecules to their ligands.

In a further aspect, the invention provides diagnostic assays for detecting diseases related to expression of the adhesin polypeptides of the present invention in infected host cells. In one embodiment, a process for diagnosing a proteobacterial infection comprises contacting a biological sample from a possibly infected subject with a labeled antibody which binds to the conserved polypeptide described herein; and measuring the signal generated by the label with a suitable assay. Detection of said signal indicates the presence of an adhesin molecule from the proteobacteria.

In another aspect, the invention provides a diagnostic reagent which comprises a composition capable of binding to a conserved proteobacterial polypeptide of the invention, the composition associated with a detectable label.

In another aspect of the present invention, there are provided antagonists,

which can be targeted against the proteobacterial adhesin to prevent its binding with the infected host=s cell surface proteins. Antagonists of adhesin binding activity can be used in the treatment of proteobacterial infection.

In another aspect, the invention provides a method for utilizing the polypeptides of the present invention for the screening of chemical or natural compounds or ligands thereof which inhibit or retard interaction of proteobacterial adhesins with other proteins, including their ligands expressed on the cells of their infected hosts. For example, in one embodiment of this aspect, the invention provides a method for identifying compounds which antagonize the binding of a proteobacterial adhesin to its ligand comprising the steps of providing a sample of the ligand or a cell which expresses the ligand immobilized on a support; contacting the sample with a known amount of a polypeptide of this invention and a known amount of a test compound; washing unbound materials from said sample; contacting the sample with a labeled reagent which binds to said polypeptide; washing unbound reagent from said sample; and measuring the amount of signal generated by said label. The amount of signal generated is inversely proportional to the ability of the test compound to disrupt or inhibit binding between said polypeptide and its ligand. Finally, the method involves identifying those test compounds as antagonists which are associated with a low signal.

In yet a further aspect, the invention provides a method for generating a small molecule which antagonizes the binding between a proteobacterial adhesin and its ligand comprising analyzing an antibody to a polypeptide of this invention in a computer modelling program.

In still another aspect, the invention provides products, compositions, processes and methods that utilize the aforementioned polypeptides and polynucleotides, antibodies and small molecules, as well as other antagonists of bacterial adhesins, for scientific research, biological, clinical and therapeutic purposes, synthesis of DNA and manufacture of DNA vectors, *inter alia*.

In yet a further aspect, the invention provides products, compositions and methods, *inter alia*, for, among other things, assessing proteobacterial infection or the expression of the bacterial adhesin in an infected host by determining the presence of the adhesins with antibodies of this invention or with nucleotide probes of this invention.

Other objects, features, advantages and aspects of the present invention will become apparent to those of skill in the art from the following description.

Breif Description of the Drawings

Figure 1 A shows a graphical illustration of the coiled-coiled probability of YadA, UspA1 and UspA2.

Figure 1 B shows a graphical illustration of normalized FFT intensity for head sequence and stalk sequence.

Figure 2 A, B and C show sequence comparisons of proteins of the invention.

Detailed Description of the Invention

The present invention meets the needs in the art by providing immunogenic sequences which are derived from the conserved, sequences in adhesins of proteobacteria, as well as fusion proteins, pharmaceutical compositions and methods of utilizing these sequences and compositions for diagnostic, therapeutic and vaccine methods and compositions.

I. Polypeptides of the Invention

The present invention relates to novel polypeptide sequences, which are isolated, highly conserved sequences of extracellular domains of certain pathogens of the family *Proteobacteria*, consensus sequences thereof, and other variants and analogs which share the immunogenic function of the native isolated sequences. The inventor identified significantly similar proteins in the beta branch of Proteobacteria, e.g., the species *Neisseria*, and in the gamma branch of Proteobacteria, e.g., the species *Actinobacillus*, *Heomophilus*, *Moraxella* and *Yersinia*. In particular, the inventor surprisingly discovered and isolated a highly conserved sequence of approximately 20 to about 24 amino acid residues, with approximately 14-18 residues centered around a conserved block of four hydrophobic residues, ending with an invariant glycine, e.g., Lys-Ala-Ala-**Gly**, which has not previously been identified in any eubacterial extracellular domain. This sequence, which in the non-fimbrial adhesin YadA of the *Yersinia* species, is found at the end of the head domain, partly overlapping the tetradecad repeat, was found in many putative open reading frames from Proteobacterial genomes. Surprisingly, the sequence was not readily

detectable by BLAST® program homology searches, and it is significantly misaligned by automated alignment programs, such as the CLUSTAL® program and the PILEUP® program.

An isolated highly conserved *Proteobacterial* sequence of this invention comprises about 20-24 amino acids. Presented herein are several generic formulae which are conserved or partially conserved sequences of the present inventions.

The N-terminal R in each sequence below may represent hydrogen (i.e., the hydrogen on the unmodified N terminal amino acid), or a lower alkyl, or a lower alkanoyl having 1 to 10 carbon atoms. R may also include a sequence of between 1 to about 25 amino acids, optionally substituted with a lower alkyl or lower alkanoyl. The C-terminal R^2 can be the hydroxyl group on the C terminal amino acid or an amide, optionally substituted with a lower alkyl or a lower alkanoyl having from 1 to 25 amino acids. It should be understood that R and R^2 will be completely omitted or defined differently, where the polypeptide or fragment of this invention is employed as part of a fusion protein with other proteins, as discussed below. For example, the polypeptides of this invention may be preceded at the N terminus (e.g., R) by a selected signal peptide and followed at the C terminus by an optional spacer sequence (e.g., R^2). These varied definitions for the N and C termini of the polypeptides of this invention are the same for all of the following formulae for embodiments of the polypeptides or fusion proteins of this invention.

Thus, in one embodiment, an isolated consensus sequence of this invention comprises a sequence of the formula:

R-Arg- X^1 - X^2 -Thr- X^3 - X^4 -Ala- X^5 - **Gly**- X^6 - X^7 - X^8 - Thr-Asp-Ala-Val-Asn- X^9 - X^{10} -Gln-Leu-R^2 [SEQ ID NO: 1].

According to this formula, X^1 can be Gln, Lys, Thr, Val, or Arg; X^2 and the hydrophobic residue X^4 are independently Leu, Ile, or Val; X^3 can be His, Gly, Ser, Asn, or Gln; the hydrophobic residue X^5 can be Ala, Lys, Val, Asp, Pro, Asn, Gly, or Glu; X^6 can be Thr, Val, Ser, Arg, Leu, Gln, Asp, Glu, Lys, or Asn; X^7 can be Lys, Glu, Ala, Gln, Ile, Asn, or Val; X^8 can be Asp, Asn, Gly, Ala, Ser, or Pro; X^9 can be Val, Leu, Phe, Gly, Lys, Met, or Ile; and X^{10} can be Ala, Gly, Ser, Asp, Arg, or Lys. Some desirable sequences of this formula include those in which X^2 is Ile and X^4 is Val.

Specifically desirable sequences of this formula are

R-Arg Gln Leu Thr His Leu Ala Ala Gly Thr Lys Asp Thr Asp Ala Val Asn Val Ala Gln Leu- R^2 [SEQ ID NO: 2]. The amino acid sequence was isolated from the YadA adhesin sequence and from the YopA preprotein sequence of *Yersinia enterocolitica* (Genbank identification numbers 401465 and 96988, respectively);

R-Arg Gln Leu Thr His Leu Ala Ala Gly Thr Glu Asp Thr Asp Ala Val Asn Val Ala Gln Leu- R^2 [SEQ ID NO: 3]. The amino acid sequence was isolated from the YadA adhesin sequences of *Yersinia pseudotuberculosis* and from *Yersinia pestis* (Genbank identification number 141104 and raw genomic sequence from the Sanger Institute, respectively);

R-Arg Gln Leu Thr Asn Ile Ala Val Gly Thr Gln Gly Thr Asp Ala Val Asn Leu Asp Gln Leu- R^2 [SEQ ID NO: 4]. The amino acid sequence was isolated from the raw genomic sequence of the YadA adhesin of *Yersinia pestis*.

Some other isolated polypeptide sequences of the invention include those sequences of the formula:

R-Arg Gln Ile Thr X^1 Val Lys X^2 Gly Val X^3 X^4 Thr Asp X^5 X^6 Asn Val X^7 Gln Leu - R^2 [SEQ ID NO: 5]. According to this formula, X^1 and X^7 are independently Gly or Ser; X^2 is Ala or Lys; X^3 is Ala or Glu; X^4 is Asp or Asn; X^5 is Ala or Thr; and X^6 is Ala or Ile. R and R^2 are as defined above. For example, specifically desirable sequences of this formula are

R-Arg Gln Ile Thr Gly Val Lys Ala **Gly** Val Ala Asp Thr Asp Ala Ala Asn Val Gly Gln Leu- R^2 [SEQ ID NO: 6]. This amino acid sequence was isolated from the raw genomic sequence (Sanger) of *Yersinia pestis;* and

R-Arg Gln Ile Thr Gly Val Lys Lys **Gly** Val Glu Asn Thr Asp Thr Ile Asn Val Ser Gln Leu- R^2 [SEQ ID NO: 7]. This amino acid sequence was isolated from raw genomic sequence (Sanger) of *Yersinia pestis*.

Other isolated polypeptide sequences of the invention include sequences of the formula:

R-Arg Lys Ile Thr Gly Val Ala Ala **Gly** Ser Ala X^1 X^2 Asp X^3 Val Asn Val Asn Gln Leu- R^2 [SEQ ID NO: 8], in which X^1 is Asp or Ser; X^2 is Tyr or Ser; and X^3 is Val or Ala. R and R^2 are as defined above. R and R^2 are as defined above. For example, specifically desirable sequences of this formula isolated from the raw genomic sequence

(Sanger) of *Yersinia pestis* are:

R-Arg Lys Ile Thr Gly Val Ala Ala **Gly** Ser Ala Asp Tyr Asp Val Val Asn Val Asn Gln Leu- R^2 [SEQ ID NO: 9];

R-Arg Lys Ile Thr Gly Val Ala Ala **Gly** Ser Ala Asp Tyr Asp Ala Val Asn Val Asn Gln Leu- R^2 [SEQ ID NO: 10]; and

R-Arg Lys Ile Thr Gly Val Ala Ala **Gly** Ser Ala Ser Ser Asp Ala Val Asn Val Asn Gln Leu- R^2 [SEQ ID NO: 11]. SEQ ID NOS: 9 and 10 were isolated as multiple copies of the conserved sequence from the same proteins.

Still other embodiments of isolated polypeptide sequences of this invention have the formula:

R-Arg Thr Val Ser Asn Val Ala Asp Gly X^1 X^2 Ala X^3 Asp Ala Val Asn Leu Arg Gln Leu- R^2 [SEQ ID NO: 12], in which X^1 is Arg or Leu; X^2 is Glu or Gln; and X^3 is Met or Thr. R and R^2 are as defined above. For example, specifically desirable sequences of this formula which were isolated from the raw genomic sequences of *Yersinia pestis* include:

R-Arg Thr Val Ser Asn Val Ala Asp Gly Arg Glu Ala Met Asp Ala Val Asn Leu Arg Gln Leu- R^2 [SEQ ID NO: 13], which was isolated from the same protein as SEQ ID NOS: 9 and 10 above; and

R-Arg Thr Val Ser Asn Val Ala Asp Gly Leu Gln Ala Thr Asp Ala Val Asn Leu Arg Gln Leu- R^2 [SEQ ID NO: 14], which was isolated from the same protein as SEQ ID NOS: 11 above.

Other examples of polypeptide sequences of this invention include the following:

R-Val Val Ile Asp Asn Val Ala Asn Gly Asp Ile Ser Ala Thr Ser Thr Asp Ala Ile Asn Gly Ser Gln Leu- R^2 [SEQ ID NO: 15]; this amino acid sequence was isolated from a Hia sequence of *Haemophilus influenzae* (Genbank identification number 1235666);

R-Val Val Ile Asp Asn Val Ala Asn Gly Glu Ile Ser Ala Thr Ser Thr Asp Ala Ile Asn Gly Ser Gln Leu- R^2 [SEQ ID NO: 16]; this amino acid sequence was isolated from the hsf gene product of *Haemophilus influenzae* (Genbank identification number 1666683);

R-Lys Arg Ile Ala Asn Val Ala Lys Gly Lys Ala Pro Thr Asp Ala Val Asn Met Ser Gln Leu- R^2 [SEQ ID NO: 17]; this amino acid sequence was isolated from the raw genomic sequence (Oklahoma University) of *Actinobacillus actinomycetemcomitans*;

R-Arg Arg Ile Ile Asn Val Ala Gly Gly Arg Asn Asp Thr Asp Ala Val Asn Ile Ala Gln Leu- R^2 [SEQ ID NO: 18]; this amino acid sequence was isolated from the raw genomic sequence (Oklahoma University) of *Actinobacillus actinomycetemcomitans*;

R-Asn Arg Ile Thr Gly Val Ala Glu Gly Thr Gln Asp Asp Asp Ala Val Asn Phe Lys Gln Leu- R^2 [SEQ ID NO: 19]; this amino acid sequence was isolated from the raw genomic sequence (Oklahoma University) of *Actinobacillus actinomycetemcomitans*;

R-Arg Gln Ile Lys Asn Val Ala Ala Gly Asn Val Ala Ala Asn Ser Thr Asp Ala Val Asn Gly Ser Gln Leu- R^2 [SEQ ID NO: 20]; this amino acid sequence was isolated from the raw genomic sequence (Oklahoma University) of *Actinobacillus actinomycetemcomitans*;

R-Lys Lys Ile Thr Asn Val Ala Asp Gly Val Ile Ala Ala Asn Ser Lys Asp Ala Val Asn Gly Gly Gln Leu- R^2 [SEQ ID NO: 21]; this amino acid sequence was isolated from the raw genomic sequence (Oklahoma University) of *Actinobacillus actinomycetemcomitans*;

R-Arg Lys Ile Val Gly Val Asp Asp Gly Val Asn Asp Phe Asp Ala Val Asn Val Arg Gln Leu- R^2 [SEQ ID NO: 22]; this amino acid sequence was isolated from the raw genomic sequence (Oklahoma University) of *Neisseria gonorrheae*;

R-Arg Gln Ile Thr Asn Val Ala Pro Ala Thr Gln Gly Thr Asp Ala Val Asn Phe Asp Gln Leu- R^2 [SEQ ID NO: 23]; this amino acid sequence was isolated from the raw genomic sequence (Sanger) of *Yersinia pestis*;

R-Arg Gln Ile Val Asn Val Gly Ala Gly Gln Ile Ser Asp Thr Ser Thr Asp Ala Val Asn Gly Ser Gln Leu- R^2 [SEQ ID NO: 24], this amino acid sequence was isolated from the high molecular weight outer membrane protein of *Moraxella catarrhalis* (Genbank identification no. 2772586); and

R-Gly Arg Ile Thr Gln Val Ala Asp Gly Val Asn Asp Lys Asp Ala Val Asn Lys Ser Gln Leu- R^2 [SEQ ID NO: 25]; this amino acid sequence was isolated from the raw genomic sequence (Oklahoma University) of *Actinobacillus actinomycetemcomitans*.

Polypeptides of the present invention also include the polypeptide of the

sequences of the invention set forth herein, as well as other polypeptides which share at least 50% identity to the consensus sequences described above and/or to the highly conserved sequences in proteobacterial extracellular domains of Proteobacterial species: *Neisseria, Actinobacillus, Haemophilus, Moraxella* and *Yersinia*, according to the algorithm BESTFIT from the GCG program package [J. Devereux et al., Nucl. Acids Res., 12(1):387 (1984)]. Other polypeptide sequences of this invention share at least 70% identity to the consensus sequences described above and/or to the highly conserved sequences in proteobacterial extracellular domains. Still other polypeptide sequences of this invention share at least 90% identity to the consensus sequences described above and/or to the highly conserved sequences in proteobacterial extracellular domains. These polypeptides are also anticipated to be useful in the compositions and methods for which the above-identified polypeptides are useful.

As known in the art, "similarity" between two polypeptides is determined by comparing the amino acid sequence and its conserved amino acid substitutes of one polypeptide to the sequence of a second polypeptide. Moreover, also known in the art is "identity" which means the degree of sequence relatedness between two polypeptide or two polynucleotide sequences as determined by the identity of the match between two lengths of such sequences. Both identity and similarity can be readily calculated [COMPUTATIONAL MOLECULAR BIOLOGY, Lesk, A.M., ed., Oxford University Press, New York, (1988); BIOCOMPUTING: INFORMATICS AND GENOME PROJECTS, Smith, D.W., ed., Academic Press, New York, (1993); COMPUTER ANALYSIS OF SEQUENCE DATA, PART I, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, (1994): SEQUENCE ANALYSIS IN MOLECULAR BIOLOGY, von Heinje, G., Academic Press, (1987); and SEQUENCE ANALYSIS PRIMER, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, (1991)]. While there exist a number of methods to measure identity and similarity between two polynucleotide or polypeptide sequences, the terms "identity" and "similarity" are well known to skilled artisans [H. Carillo and D. Lipton, SIAM J. Applied Math., 48:1073 (1988)]. Methods commonly employed to determine identity or similarity between two sequences include, but are not limited to, those disclosed in Guide to Huge Computers, Martin J. Bishop, ed., Academic Press, San Diego, 1994, and H. Carillo and D. Lipton,

SIAM J. Applied Math., 48:1073 (1988). Preferred methods to determine identity are designed to give the largest match between the two sequences tested. Methods to determine identity and similarity are codified in computer programs. Preferred computer program methods to determine identity and similarity between two sequences include, but are not limited to, GCG program package [J. Devereux et al., Nucl. Acids Res., 12(1):387 (1984)], BLAST [S. F. Atschul et al., J. Mol. Biol., 215:403 (1990)] and FASTA (Pearson) programs. For instance, searches for sequence similarities in databases with other proteobacterial species are likely to detect other similar highly conserved sequences.

In general, as used herein, the term Apolypeptide≡ encompasses the above-identified polypeptides, and all modifications, particularly those that are present in polypeptides synthesized by expressing a polynucleotide in a host cell. In one embodiment, the polypeptides of the present invention are preferably provided in an isolated form, and preferably are purified to homogeneity. "Isolated" means altered "by the hand of man" from its natural state; i.e., that, if it occurs in nature, it has been changed or removed from its original environment, or both. For example, a naturally occurring polypeptide naturally present in a living animal in its natural state is not "isolated", but the same polypeptide separated from the coexisting materials of its natural state is "isolated", as the term is employed herein. Similarly, the polypeptides may occur in a composition, such as a media, formulations, solutions for introduction of polypeptides, for example, into cells, compositions or solutions for chemical or enzymatic reactions, for instance, which are not naturally occurring compositions, and, therein remain isolated polypeptides within the meaning of that term as it is employed herein.

In another embodiment, the polypeptide of the present invention may be a recombinant polypeptide expressed from a prokaryotic or eukaryotic host, including, for example, bacterial, yeast, higher plant, insect and mammalian cells. In still another embodiment, a polypeptide of the invention may be a synthetic polypeptide.

The invention also relates to variants, analogs, derivatives and fragments of these isolated or consensus polypeptides, and variants, analogs and derivatives of the fragments. "Variant(s)" of polypeptides, as the term is used herein, are polypeptides that differ in amino acid sequence from the above-identified polypeptides, which serve as reference polypeptides. Generally, differences are limited so that the sequences of the

reference and the variant are closely similar overall and, in many regions, identical. A variant and reference polypeptide may differ in amino acid sequence by one or more substitutions, additions, deletions, fusions and truncations, which may be present in any combination. Among preferred variants are those that vary from a reference by conservative amino acid substitutions. Such substitutions are those that substitute a given amino acid in a polypeptide by another amino acid of like characteristics. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu and Ile; interchange of the hydroxyl residues Ser and Thr, exchange of the acidic residues Asp and Glu, substitution between the amide residues Asn and Gln, exchange of the basic residues Lys and Arg and replacements among the aromatic residues Phe and Tyr.

The terms "fragment," "derivative" and "analog" when referring to the polypeptide of the sequences of the invention set forth herein, means a polypeptide which retains essentially the same biological function or activity as such polypeptide, i.e., functions as an immunogen or retains the ability to bind its ligand expressed on a host cell. The fragment, derivative or analog of the polypeptide the sequences of the invention set forth herein may be (i) one in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue) and such substituted amino acid residue may or may not be one encoded by the genetic code; (ii) one in which one or more of the amino acid residues includes a substituent group; (iii) one in which the polypeptide is fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol); or (iv) one in which the additional amino acids are fused to the polypeptide, such as a leader or secretory sequence or a sequence which is employed for purification of the polypeptide. Such fragments, derivatives and analogs are deemed to be within the scope of those skilled in the art from the teachings herein.

Further, particularly preferred in this regard are variants, analogs, derivatives and fragments having the amino acid sequence of an above-described polypeptide of the sequences of the invention set forth herein, in which several, a few, 5 to 10, 1 to 5, 1 to 3, 2, 1 or no amino acid residues are substituted, deleted or added, in any combination. Especially preferred among these are silent substitutions, additions and

deletions, which do not alter the properties and activities of the polypeptide of the invention. Most highly preferred are polypeptides having the amino acid sequence of the sequences of the invention set forth herein without substitutions.

Additionally, shorter versions of the above-identified sequences and formulae may also be useful where these fragments retain immunogenicity. It is anticipated that for the specific polypeptide identified above, an optional truncation of 1 amino acid at the C terminus is likely to produce a useful fragment. Additionally, a truncation of up to about 4 or 5 amino acids from the N terminus of the polypeptides of this invention is also likely to produce a useful peptide fragment. Thus, in one embodiment a polypeptide has a sequence similarity or identity of at least about 80% to the sequence of the sequences of the invention set forth herein, and more preferably at least 90% similarity (more preferably at least 95% identity) to a polypeptide of the invention, and still more preferably at least 95% similarity (still more preferably at least 95% identity) to the polypeptide of the sequences of the invention set forth herein and also include portions of such polypeptides with such portion of the polypeptide generally containing at least 14 amino acids. Other embodiments of fragments of this invention contain at least 18 amino acids of the highly conserved or consensus sequences; and still others contain at least 20 amino acids.

Fragments or portions of the polypeptides of the present invention may be employed for producing the corresponding full-length polypeptide by peptide synthesis; therefore, the fragments may be employed as intermediates for producing the full-length polypeptides. Fragments or portions of the polynucleotides of the present invention may be used to synthesize full-length polynucleotides of the present invention. Fragments may be "free-standing," i.e., not part of or fused to other amino acids or polypeptides, or they may be comprised within a larger polypeptide of which they form a part or region. When comprised within a larger polypeptide, the presently discussed fragments most preferably form a single continuous region. However, several fragments may be comprised within a single larger polypeptide. For instance, certain preferred embodiments relate to a fragment of a polypeptide of the present invention comprised within a precursor polypeptide designed for expression in a host cell and having heterologous pre- and pro-polypeptide regions fused to the amino terminus of the polypeptide or fragment and an additional

region fused to the carboxyl terminus of the fragment. Therefore, fragments in one aspect of the meaning intended herein, refers to the portion or portions of a fusion polypeptide or fusion protein derived from the highly conserved Proteobacterial sequences or the consensus sequence therefrom.

As representative examples of polypeptide fragments of the invention, there may be mentioned those which have from about 5-15, 10-20, and 24 amino acids in length. In this context "about" includes the particularly recited range and ranges larger or smaller by several, a few, 5, 4, 3, 2 or 1 amino acid residues. For instance, about 20 amino acids in this context means a polypeptide fragment of 20 plus or minus several, a few, 5, 4, 3, 2 or 1 amino acid residues. Further preferred fragments are those that have a chemical, biological or other activity of a conserved polypeptide of the invention, including those with a similar activity or an improved activity, or with a decreased undesirable activity. Highly preferred in this regard are the recited ranges plus or minus as many as 5 amino acids at either or at both extremes. Particularly highly preferred are the recited ranges plus or minus as many as 3 amino acids at either or at both the recited extremes. Especially particularly highly preferred are ranges plus or minus 1 amino acid at either or at both extremes or the recited ranges with no additions or deletions.

Polypeptides of this invention may also contain amino acids other than the 20 amino acids commonly referred to as the 20 naturally occurring amino acids, and that these amino acids, including the terminal amino acids, may be modified, either by natural processes, such as processing and other post-translational modifications, or by chemical modification techniques which are well known to the art. The numerous common modifications that occur naturally in polypeptides are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature, and are well known to those of skill in the art.

Among the known modifications which may be present in polypeptides of the present invention are, without limitation, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of

cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination.

Such modifications are well known to those of skill and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as, PROTEINS - STRUCTURE AND MOLECULAR PROPERTIES, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York, 1993. Many detailed reviews are available on this subject, such as those provided by Wold, F., "Posttranslational Protein Modifications: Perspectives and Prospects", pgs. 1-12 in POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS, B. C. Johnson, Ed., Academic Press, New York, 1983; Seifter et al., Meth. Enzymol., 182:626-646 (1990), and Rattan et al., Ann. N.Y. Acad. Sci., 663:48-62 (1992).

The polypeptides of this invention may be linear, or branched as a result of ubiquitination, or they may be circular, with or without branching, generally as a result of posttranslational events, including natural processing event and events brought about by human manipulation which do not occur naturally. Circular, branched and branched circular polypeptides may be synthesized by non-translation natural processes and by entirely synthetic methods, as well.

Modifications can occur anywhere in a polypeptide of this invention, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. In fact, blockage of the amino or carboxyl group in a polypeptide, or both, by a covalent modification, is common in naturally occurring and synthetic polypeptides and such modifications may be present in polypeptides of the present invention. For instance, the amino terminal residue of polypeptides made in *E. coli*, prior to processing, almost invariably will be N-formylmethionine. The modifications that occur in a polypeptide often will be a function of how it is produced. For polypeptides produced by expression in a host, for instance, the nature and extent of the modifications in large part will be

determined by the host cell's posttranslational modification capacity and the modification signals present in the polypeptide amino acid sequence. For instance, as is well known, glycosylation often does not occur in bacterial hosts such as *E. coli*. Accordingly, when glycosylation is desired, a polypeptide should be expressed in a glycosylating host, generally a eukaryotic cell. Insect cells often carry out the same posttranslational glycosylations as mammalian cells and, for this reason, insect cell expression systems have been developed to express efficiently mammalian proteins having the native patterns of glycosylation, *inter alia*. Similar considerations apply to other modifications.

It will be appreciated that the same type of modification may be present in the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications.

Several of the proteins were identified by this inventor as containing multiple copies of a conserved sequence, most notably a protein in contig 763 of the *Actinobacillus actinomyetemcomitans* genome with four copies. The conserved sequence (or in proteins with multiple copies, the last copy) was found at the N-terminal end of a predicted coiled-coil rod, generally corresponding to the left-handed coiled-coil segment of the YadA rod. Occasionally, as in YadA, it is extended by additional coiled-coil sequences. Such coiled coils are highly immunogenic, yet have the ability to withstand rapid mutation without losing their structure. They therefore provide bacteria with excellent, evolving decoys against the host immune system. One function of the coiled-coil rod in the YadA class of proteins is to distract the host immune system from the highly conserved (and functionally essential) sequence.

All proteins in the YadA class that are not visibly incomplete contain a C terminal transmembrane anchor, preceded by a conserved region of heptad repeat, which is in turn preceded (except in *Moraxella* A2) by the conserved sequence, suggesting a common architecture. The head domain repeats of Yad A are not recognizable in a number of these proteins even though the conserved sequence appears to be an extension of these repeats, suggesting that it may have evolved form them but developed an independent functionality.

The highly conserved sequence of the present invention acts as a binding motif, and is likely involved in host cell recognition, auto-agglutination, or cell defense

through binding of complement inhibitor factor H. It may bind collagen, although a truncated mutant was unable to bind collagen but still contained this sequence. These polypeptides of this invention or immunogenic fragments thereof are useful in the development of antibodies and the design of diagnostic probes, screening methods for the development of vaccine agents to prevent bacterial infection, and the like discussed below.

II. Fusion Proteins or Multiple Antigenic Complexes of The Invention

A "fusion protein" as the term is used herein, is a protein encoded by a polynucleotide sequence encoding a polypeptide, variant, or fragment of this invention to another, often unrelated, gene or fragments thereof. The Aother≡ protein to which the polypeptide of this invention is fused or coupled may be selected from among any proteins or peptides which are at least 90% likely to form a coiled coil, as defined by the COILS algorithm [A. Lupas *et al*, Science, 252:1162-1164 (1991), incorporated by reference herein]. See, also, European Patent Application No. EP-A-0 464 533 [Canadian counterpart Patent Application No. 2045869] which discloses fusion proteins comprising various portions of constant regions of immunoglobulin molecules together with another human protein or part thereof. In many cases, employing an immunoglobulin Fc region as a part of a fusion protein is advantageous for use in therapy and diagnosis resulting in, for example, improved pharmacokinetic properties [See, e.g., European Patent Application No. EP-A 0 232 262]. For some uses, it would be desirable to be able to delete the Fc part after the fusion protein has been expressed, detected and purified. Accordingly, it may be desirable to link the two components of the fusion protein with a chemically or enzymatically cleavable linking region. This is the case when the Fc portion proves to be a hindrance to use in therapy and diagnosis, for example, when the fusion protein is to be used as an antigen for immunizations. In drug discovery, for example, human proteins have been fused with Fc portions for use in high-throughput screening assays to identify antagonists of those proteins. See, D. Bennett et al., J. Mol. Recog., 8:52-58 (1995); and K. Johanson et al., J. Biol. Chem., 270(16):9459-9471 (1995).

Thus, this invention also relates to genetically engineered fusion proteins comprised of one of the conserved proteobacterial sequences or a variant, derivative or

fragment thereof, and of various portions of the constant regions of heavy or light chains of immunoglobulins of various subclasses (IgG, IgM, IgA, IgE). Preferred as an immunoglobulin is the constant part of the heavy chain of human IgG, particularly IgG1, where fusion takes place at the hinge region. In a particular embodiment, the Fc part can be removed simply by incorporation of a cleavage sequence which can be cleaved with blood clotting factor Xa.

Membrane-bound receptors are particularly useful in the formation of fusion proteins. Such receptors are generally characterized as possessing three distinct structural regions: an extracellular domain, a transmembrane domain and a cytoplasmic domain. This invention contemplates the use of one or more of these regions as components of a fusion protein. Examples of such fusion protein technology can be found in International Patent Application Nos. WO94/29458 and WO94/22914.

The fusion proteins of the present invention may be prepared and used in a variety of forms, for example, chemically synthesized or as recombinant peptides, polypeptides, proteins, fusion proteins or fused peptides. As one embodiment, a composition of the present invention may be a synthetic peptide, containing single or multiple copies of the same or different polypeptide of this invention, coupled to a selected carrier protein. In this embodiment of a composition of this invention, one or more polypeptides or fragments thereof as described above may be coupled or fused to a carrier protein, or several may be admixed to create a immunogenic composition.

For this embodiment, the carrier protein is desirably a protein or other molecule which can enhance the immunogenicity of the selected immunogen. Such a carrier may be a larger molecule which has an adjuvanting effect. Exemplary conventional protein carriers include, without limitation, *E. coli* DnaK protein, galactokinase (galK, which catalyzes the first step of galactose metabolism in bacteria), ubiquitin, α-mating factor, ß-galactosidase, and influenza NS-1 protein. Toxoids (i.e., the sequence which encodes the naturally occurring toxin, with sufficient modifications to eliminate its toxic activity) such as diphtheria toxoid and tetanus toxoid may also be employed as carriers. Similarly a variety of bacterial heat shock proteins, e.g., mycobacterial hsp-70 may be used. Glutathione reductase (GST) is another useful carrier. One of skill in the art can readily select an appropriate carrier.

A polypeptide or fusion protein of the present invention may also be modified to increase its immunogenicity. For example, the polypeptide or fusion protein may be coupled to chemical compounds or immunogenic carriers, provided that the coupling does not interfere with the desired biological activity of either the polypeptide or the carrier. For a review of some general considerations in coupling strategies, see Antibodies. A Laboratory Manual. Cold Spring Harbor Laboratory. ed. E. Harlow and D.Lane (1988). Useful immunogenic carriers known in the art, include, without limitation, keyhole limpet hemocyanin (KLH); bovine serum albumin (BSA), ovalbumin, PPD (purified protein derivative of tuberculin); red blood cells; tetanus toxoid; cholera toxoid; agarose beads; activated carbon; or bentonite. Useful chemical compounds for coupling include, without limitation, dinitrophenol groups and arsonilic acid. The polypeptide or fusion protein antigen may also be modified by other techniques, such as denaturation with heat and/or SDS.

In particularly desirable immunogen-carrier protein construct, one or more polypeptides or fragments of this invention may be covalently linked to a mycobacterial or *E. coli* heat shock protein 70 (hsp70) [K. Suzue et al, J. Immunol., 156:873 (1996)]. In another desirable embodiment, the composition is formed by covalently linking the polypeptide sequence(s) to diphtheria toxoid.

Alternatively, the polypeptides are assembled as multi-antigenic peptide (MAP) complexes [see, e.g., European Patent Application 0339695, published November 2, 1989] or as simple mixtures of antigenic proteins/peptides and employed to elicit high titer antibodies capable of binding the selected antigen(s) as it appears in the biological fluids of an infected animal or human.

In any of the above-mentioned fusion protein or MAP compositions, each amino acid sequence may be optionally separated by optional amino acid sequences called "spacers". Spacers are sequences of between 1 to about 4 amino acids which are interposed between two sequences to permit linkage therebetween without adversely effecting the three dimensional structure of the fusion protein. Spacers may also contain restriction endonuclease cleavage sites to enable separation of the sequences, where desired. Suitable spacers or linkers are known and may be readily designed and selected by one of skill in the art. This invention also relates to processes for the preparation of

these fusion proteins by genetic engineering, and to the use thereof for diagnosis and therapy.

III. Polynucleotide Sequences of this Invention

The present invention provides an isolated nucleic acid (polynucleotide) which encodes the highly conserved isolated polypeptide sequences having the amino acid sequences defined above [the sequences of the invention set forth herein].

With respect to polynucleotides, the term "isolated" means that it is separated from the chromosome and cell in which it naturally occurs. As part of or following isolation, such polynucleotides can be joined to other polynucleotides, such as DNAs, for mutagenesis, to form fusion proteins, and for propagation or expression in a host, for instance. The isolated polynucleotides, alone or joined to other polynucleotides such as vectors, can be introduced into host cells, in culture or in whole organisms. Introduced into host cells in culture or in whole organisms, such DNAs still would be isolated, because they would not be in their naturally occurring form or environment. The term Apolynucleotide(s)" generally refers to any polyribonucleotide or polydeoxyribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA which encodes a polypeptide of the present invention. The term includes only coding sequence for the polypeptide as well as a polynucleotide which includes additional coding and/or non-coding sequence. The term also encompasses polynucleotides that include a single continuous region encoding the polypeptide together with additional regions, that also may contain coding and/or non-coding sequences.

Such sequences include mRNAs, DNAs, cDNAs, genomic DNAs and fragments thereof. The polynucleotides may be single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is a mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. Single-stranded DNA may be the coding strand, also known as the sense strand, or it may be the non-coding strand, also referred to as the anti-sense strand. In addition, polynucleotide as used herein refers to triple-stranded regions comprising RNA or DNA or both RNA and DNA. The strands in

such regions may be from the same molecule or from different molecules. The regions may include all of one or more of the molecules, but more typically involve only a region of some of the molecules. One of the molecules of a triple-helical region often is an oligonucleotide.

As used herein, the term polynucleotide includes DNAs or RNAs as described above that contain one or more modified bases. Thus, DNAs or RNAs with backbones modified for stability or for other reasons are polynucleotides as that term is intended herein. Moreover, DNAs or RNAs comprising unusual bases, such as inosine, or modified bases, such as tritylated bases, to name just two examples, are polynucleotides as the term is used herein. It will be appreciated that a great variety of modifications have been made to DNA and RNA that serve many useful purposes known to those of skill in the art. The term polynucleotide, as it is employed herein, embraces such chemically, enzymatically or metabolically modified forms of polynucleotides, as well as the chemical forms of DNA and RNA characteristic of viruses and cells, including *inter alia* simple and complex cells.

The invention also relates to, among others, polynucleotides encoding the aforementioned polypeptide fragments, polynucleotides that hybridize to polynucleotides encoding the fragments, particularly those that hybridize under stringent conditions, and polynucleotides, such as PCR primers, for amplifying polynucleotides that encode the fragments. In these regards, preferred polynucleotides are those that correspond to the preferred fragments, as discussed above.

The sequences which encode the desired highly conserved or consensus polypeptide as defined above include polynucleotides with a different coding sequence, which, as a result of the redundancy (degeneracy) of the genetic code, encode the same polypeptide or desired fragment thereof of any of the sequences of the invention set forth herein. Among the particularly preferred embodiments of this aspect of the invention are naturally occurring alleles of the bacterial adhesins which contain the conserved sequences described herein as well as analogs and biologically active and diagnostically or therapeutically useful variants, derivatives, and fragments thereof.

Polynucleotides of the present invention which encode the polypeptide of this invention may include, but are not limited to, the coding sequence for the polypeptide,

by itself; the coding sequence for the polypeptide and additional coding sequences, such as transcribed, non-translated sequences that play a role in transcription, and mRNA processing, including splicing and polyadenylation signals, for example, for ribosome binding and stability of mRNA. Coding sequences which provide additional functionalities may also be incorporated into the polypeptide.

"Variant(s)" of polynucleotides, as the term is used herein, are polynucleotides that differ in nucleotide sequence from a reference polynucleotide. Generally, differences are limited so that the nucleotide sequences of the reference and the variant are closely similar overall and, in many regions, identical. Changes in the nucleotide sequence of the variant may be silent. That is, they may not alter the amino acids encoded by the polynucleotide. Where alterations are limited to silent changes of this type, a variant will encode a polypeptide with the same amino acid sequence as the reference. Also as noted below, changes in the nucleotide sequence of the variant may alter the amino acid sequence of a polypeptide encoded by the reference polynucleotide. Such nucleotide changes may result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence, as discussed below. The present invention further relates to variants of the herein above-described polynucleotides which encode for fragments, analogs and derivatives of the polypeptides of this invention. A variant of the polynucleotide may be a naturally occurring polynucleotide that encodes the sequences of the invention set forth herein, such as a naturally occurring allelic variant, a variant which occurs in another Proteobacterial species, or a variant that is not known to occur naturally. As known in the art, an allelic variant is an alternate form of a polynucleotide sequence which may have a substitution, deletion or addition of one or more nucleotides, which does not substantially alter the function of the encoded polypeptide. Non-naturally occurring variants of the polynucleotide may be prepared by mutagenesis techniques, including those applied to polynucleotides, cells or organisms.

Among variants in this regard are variants that differ from the aforementioned polynucleotides by nucleotide substitutions, deletions or additions. The substitutions, deletions or additions may involve one or more nucleotides. The variants may be altered in coding or non-coding regions or both. Alterations in the coding regions

may produce conservative or non-conservative amino acid substitutions, deletions or additions.

Among the particularly preferred embodiments of the invention in this regard are polynucleotides encoding polypeptides having the amino acid sequence of the sequences of the invention set forth herein, variants, analogs, derivatives and fragments thereof, and fragments of the variants, analogs and derivatives as described above.

Using the information provided herein and available in the art, such as the polynucleotide sequences set out above, a polynucleotide of the present invention encoding a highly conserved Proteobacterial sequence may be obtained using standard cloning and screening procedures. Alternatively, the polynucleotide sequences of this invention may be produced by conventional synthetic means or a combination of both techniques.

The present invention also includes polynucleotides, wherein the coding sequence for the isolated polypeptide may be fused in the same reading frame to a polynucleotide sequence which aids in expression and secretion of a polypeptide from a host cell or which aids in the stability of the polypeptide in a cell, for example, a leader sequence which functions as a secretory sequence for controlling transport of a polypeptide from the cell. Thus, for instance, the polypeptide may be fused in frame to a marker sequence, such as a peptide, which facilitates purification of the fused polypeptide. In certain preferred embodiments of this aspect of the invention, the marker sequence is a hexa-histidine peptide, such as the tag provided in the pQE-9 vector (Qiagen, Inc.) to provide for purification of the polypeptide fused to the marker in the case of a bacterial host. Or, for example, as described in Gentz et al., Proc. Natl. Acad. Sci., USA, 1989, 86:821-824, hexa-histidine provides for convenient purification of the fusion protein. In other embodiments, the marker sequence is a hemagglutinin (HA) tag, particularly when a mammalian host, e.g. COS-7 cells, is used. The HA tag corresponds to an epitope derived from influenza hemagglutinin protein, which has been described by Wilson *et al.*, Cell, 1984, 37:767, for instance. Many other such tags are commercially available.

As discussed additionally herein regarding polynucleotide assays of the invention, polynucleotides of the invention may be used as hybridization probes for isolating other highly conserved sequences from other Proteobacterial species, or to isolate

DNA of other genes that have a high sequence similarity to the highly conserved or consensus sequences of this invention, and/or similar biological activity. Similarly such sequences may be used in computer programs to locate other similar sequences as they are reported in the databases. Such probes generally will comprise at least 15 nucleotides.

An example of such a screen comprises labeling an oligonucleotides having a sequence complementary to that of the sequence of the present invention and using it as a probe for hybridize to sequences in a library of Proteobacterial species DNA, or mRNA to determine novel conserved sequences. The polynucleotides which hybridize to the herein above-described polynucleotides in a preferred embodiment encode polypeptides which either retain substantially the same biological function or activity as the isolated or consensus polypeptide encoded by the sequences of the invention set forth herein.

The polynucleotides and polypeptides of the present invention may be employed as research reagents and materials for discovery of compositions, diagnostic methods, treatments and vaccine methods for the diagnosis, treatment or prevention of Proteobacterial infections.

IV. Vectors, Host Cells, Expression

The present invention also relates to vectors which include polynucleotides of the present invention, host cells which are genetically engineered with vectors of the invention and the production of polypeptides of the invention by recombinant techniques.

Host cells can be genetically engineered to incorporate polynucleotides and express polypeptides of the present invention. For instance, polynucleotides may be introduced into host cells using well known techniques of infection, transduction, transfection, transvection and transformation. Unless otherwise stated, transformation was performed as described in the method of Graham, F. and Van der Bb, A., Virology, 52:456-457 (1973).

The polynucleotides may be introduced alone or with other polynucleotides. Such other polynucleotides may be introduced independently, co-introduced or introduced joined to the polynucleotides of the invention. Thus, for instance, polynucleotides of the invention may be transfected into host cells with another, separate polynucleotide encoding a selectable marker, using standard techniques for co-transfection and selection

in, for instance, mammalian cells. In this case the polynucleotides generally will be stably incorporated into the host cell genome.

Alternatively, the polynucleotides may be joined to a vector or plasmid containing a selectable marker for propagation in a host. The vector construct may be introduced into host cells by the aforementioned techniques. "Plasmids" are genetic elements that are stably inherited without being a part of the chromosome of their host cell. They may be comprised of DNA or RNA and may be linear or circular. They can also encode genes that confer resistance to antibiotics. Plasmids are widely used in molecular biology as vectors used to clone and express recombinant genes. Plasmids generally are designated herein by a lower case p preceded and/or followed by capital letters and/or numbers, in accordance with standard naming conventions that are familiar to those of skill in the art. Many plasmids and other cloning and expression vectors that can be used in accordance with the present invention are well known and readily available to those of skill in the art. Moreover, those of skill readily may construct any number of other plasmids suitable for use in the invention. The properties, construction and use of such plasmids, as well as other vectors, in the present invention will be readily apparent to those of skill from the present disclosure.

Generally, a plasmid vector is introduced as DNA in a precipitate, such as a calcium phosphate precipitate, or in a complex with a charged lipid. Electroporation may also be used to introduce polynucleotides into a host. If the vector is a virus, it may be packaged *in vitro* or introduced into a packaging cell and the packaged virus may be transduced into cells. A wide variety of techniques suitable for making polynucleotides and for introducing polynucleotides into cells in accordance with this aspect of the invention are well known and routine to those of skill in the art. Such techniques are reviewed at length in Sambrook et al, MOLECULAR CLONING, A LABORATORY MANUAL, 2nd Ed.; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989, which is illustrative of the many laboratory manuals that detail these techniques.

In accordance with this aspect of the invention the vector may be, for example, a plasmid vector, a single or double-stranded phage vector, or a single or double-stranded RNA or DNA viral vector. Such vectors may be introduced into cells as

polynucleotides, preferably DNA, by well known techniques for introducing DNA and RNA into cells. The vectors, in the case of phage and viral vectors may also be and preferably are introduced into cells as packaged or encapsidated virus by well known techniques for infection and transduction. Viral vectors may be replication competent or replication defective. In the latter case, viral propagation generally will occur only in complementing host cells.

Preferred among vectors, in certain respects, are those for expression of polynucleotides and polypeptides of the present invention. Generally, such vectors comprise cis-acting control regions effective for expression in a host operatively linked to the polynucleotide to be expressed. Appropriate trans-acting factors are either supplied by the host, supplied by a complementing vector or supplied by the vector itself upon introduction into the host.

In certain preferred embodiments in this regard, the vectors provide for specific expression. Such specific expression may be inducible expression or expression only in certain types of cells or both inducible and cell-specific expression. Particularly preferred among inducible vectors are vectors that can be induced for expression by environmental factors that are easy to manipulate, such as temperature and nutrient additives. A variety of vectors suitable to this aspect of the invention, including constitutive and inducible expression vectors for use in prokaryotic and eukaryotic hosts, are well known and employed routinely by those of skill in the art. Presently prokaryotic expression systems are preferred.

The engineered host cells can be cultured in conventional nutrient media, which may be modified as appropriate for, *inter alia*, activating promoters, selecting transformants or amplifying genes. Culture conditions, such as temperature, pH and the like, previously used with the host cell selected for expression, generally will be suitable for expression of polypeptides of the present invention as will be apparent to those of skill in the art. A great variety of expression vectors can be used to express a polypeptide of the invention. Such vectors include chromosomal, episomal and virus-derived vectors e.g., vectors derived from bacterial plasmids, bacteriophages, yeast episomes, yeast chromosomal elements, and viruses such as baculoviruses, papova viruses, SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, and vectors

derived from combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, cosmids and phagemids. Generally, any vector suitable to maintain, propagate or express polynucleotides to produce a polypeptide in a host may be used for expression in this regard.

The appropriate DNA sequence may be inserted into the vector by any of a variety of well-known and routine techniques. In general, a DNA sequence for expression is joined to an expression vector by cleaving the DNA sequence and the expression vector with one or more restriction endonucleases and then joining the restriction fragments together using T4 DNA ligase. Procedures for restriction and ligation that can be used to this end are well known and routine to those of skill. Suitable procedures in this regard, and for constructing expression vectors using alternative techniques, which also are well known and routine to those skilled in the art, are set forth in great detail in Sambrook et al.

The DNA sequence in the expression vector is operatively linked to appropriate expression control sequence(s), including, for instance, a promoter to direct mRNA transcription. Representatives of such promoters include the phage lambda PL promoter, the *E. coli* lac, trp and tac promoters, the SV40 early and late promoters and promoters of retroviral LTRs, to name just a few of the well-known promoters. It will be understood that numerous other promoters useful in this aspect of the invention are well known and may be routinely employed by those of skill in the manner illustrated by the discussion and the examples herein.

In general, expression constructs will contain sites for transcription initiation and termination, and, in the transcribed region, a ribosome binding site for translation. The coding portion of the mature transcripts expressed by the constructs will include a translation initiating AUG at the beginning and a termination codon appropriately positioned at the end of the polypeptide to be translated.

In addition, the constructs may contain control regions that regulate as well as engender expression. Generally, in accordance with many commonly practiced procedures, such regions will operate by controlling transcription. Examples include repressor binding sites and enhancers, among others. Vectors for propagation and expression generally will include selectable markers. Selectable marker genes provide a phenotypic trait for selection of transformed host cells. Preferred markers include, but are

not limited to, dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, and tetracycline or ampicillin resistance genes for culturing *E. coli* and other bacteria. Such markers may also be suitable for amplification. Alternatively, the vectors may contain additional markers for this purpose.

The vector containing the appropriate DNA sequence as described elsewhere herein, as well as an appropriate promoter, and other appropriate control sequences, may be introduced into an appropriate host using a variety of well known techniques suitable for expression therein of a desired polypeptide. Representative examples of appropriate hosts include bacterial cells, such as *E. coli*, *Streptomyces* and *Salmonella typhimurium* cells; fungal cells, such as yeast cells; insect cells such as Drosophila S2 and Spodoptera Sf9 cells; animal cells such as CHO, COS and Bowes melanoma cells; and plant cells. Host cells for a great variety of expression constructs are well known, and those of skill will be enabled by the present disclosure to routinely select a host for expressing a polypeptide in accordance with this aspect of the present invention. Presently, prokaryotic expression systems and host cells are preferred.

More particularly, the present invention also includes recombinant constructs, such as expression constructs, comprising one or more of the sequences described above. The constructs comprise a vector, such as a plasmid or viral vector, into which such a sequence of the invention has been inserted. The sequence may be inserted in a forward or reverse orientation. In certain preferred embodiments in this regard, the construct further comprises regulatory sequences, including, for example, a promoter, operably linked to the sequence. Large numbers of suitable vectors and promoters are known to those of skill in the art, and there are many commercially available vectors suitable for use in the present invention.

The following vectors, which are commercially available, are provided by way of example. Among vectors preferred for use in bacteria are pQE70, pQE60 and pQE-9, available from Qiagen; pBS vectors, Phagescript vectors, Bluescript vectors, pNH8A, pNH16a, pNH18A, pNH46A, available from Stratagene; and ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 available from Pharmacia. Among preferred eukaryotic vectors are pWLNEO, pSV2CAT, pOG44, pXT1 and pSG available from Stratagene; and pSVK3, pBPV, pMSG and pSVL available from Pharmacia. These vectors are listed

solely by way of illustration of the many commercially available and well known vectors that are available to those of skill in the art for use in accordance with this aspect of the present invention. It will be appreciated that any other plasmid or vector suitable for, for example, introduction, maintenance, propagation or expression of a polynucleotide or polypeptide of the invention in a host may be used in this aspect of the invention.

Promoter regions can be selected from any desired gene using vectors that contain a reporter transcription unit lacking a promoter region, such as a chloramphenicol acetyl transferase ("CAT") transcription unit, downstream of a restriction site or sites for introducing a candidate promoter fragment; i.e., a fragment that may contain a promoter. As is well known, introduction into the vector of a promoter-containing fragment at the restriction site upstream of the CAT gene engenders production of CAT activity, which can be detected by standard CAT assays. Vectors suitable to this end are well known and readily available. Two examples of such vectors include pKK232-8 and pCM7. Thus, promoters for expression of polynucleotides of the present invention include not only well known and readily available promoters, but also promoters that may be readily obtained by the foregoing technique, using a reporter gene.

Among known bacterial promoters suitable for expression of polynucleotides and polypeptides in accordance with the present invention are the *E. coli* lacI and lacZ promoters, the T3 and T7 promoters, the gpt promoter, the lambda PR, PL promoters and the trp promoter. Among known eukaryotic promoters suitable in this regard are the CMV immediate early promoter, the HSV thymidine kinase promoter, the early and late SV40 promoters, the promoters of retroviral LTRs, such as those of the Rous Sarcoma Virus ("RSV"), and metallothionein promoters, such as the mouse metallothionein-I promoter. Selection of appropriate vectors and promoters for expression in a host cell is a well known procedure and the requisite techniques for construction of expression vectors, introduction of the vector into the host and expression in the host are routine skills in the art.

The present invention also relates to host cells containing the above-described constructs. The host cell can be a higher eukaryotic cell, such as a mammalian cell, a lower eukaryotic cell, such as a yeast cell, or a prokaryotic cell, such as a bacterial cell. Introduction of the construct into the host cell can be effected by calcium

phosphate transfection, DEAE-dextran mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection or other methods. Such methods are described in many standard laboratory manuals, such as Davis et al. BASIC METHODS IN MOLECULAR BIOLOGY, (1986).

Constructs in host cells can be used in a conventional manner to produce the gene product encoded by the recombinant sequence. Alternatively, the polypeptides of the invention can be synthetically produced by conventional peptide synthesizers.

Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook et al.

Generally, recombinant expression vectors will include origins of replication, a promoter derived from a highly-expressed gene to direct transcription of a downstream structural sequence, and a selectable marker to permit isolation of vector containing cells following exposure to the vector. Among suitable promoters are those derived from the genes that encode glycolytic enzymes such as 3-phosphoglycerate kinase ("PGK"), α-factor, acid phosphatase, and heat shock proteins, among others. Selectable markers include the ampicillin resistance gene of *E. coli* and the trp1 gene of *S. cerevisiae*.

Transcription of DNA encoding the polypeptides of the present invention by higher eukaryotes may be increased by inserting an enhancer sequence into the vector. Enhancers are cis-acting elements of DNA, usually from about 10 to 300 bp, that act to increase transcriptional activity of a promoter in a given host cell-type. Examples of enhancers include the SV40 enhancer, which is located on the late side of the replication origin at bp 100 to 270, the cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers.

A polynucleotide of the invention encoding the heterologous structural sequence of a polypeptide of the invention generally will be inserted into the vector using standard techniques so that it is operably linked to the promoter for expression. The polynucleotide will be positioned so that the transcription start site is located appropriately

5' to a ribosome binding site. The ribosome binding site will be 5' to the AUG that initiates translation of the polypeptide to be expressed. Generally, there will be no other open reading frames that begin with an initiation codon, usually AUG, and lie between the ribosome binding site and the initiation codon. Also, generally, there will be a translation stop codon at the end of the polypeptide and a polyadenylation signal and transcription termination signal appropriately disposed at the 3' end of the transcribed region.

Appropriate secretion signals may be incorporated into the expressed polypeptide for secretion of the translated protein into the lumen of the endoplasmic reticulum, the periplasmic space or the extracellular environment. The signals may be endogenous to the polypeptide or heterologous.

The polypeptide may be expressed in a modified form, such as a fusion protein, and may include not only secretion signals but also additional heterologous functional regions. Thus, for example, a region of additional amino acids, particularly charged amino acids, may be added to the N-terminus of the polypeptide to improve stability and persistence in the host cell during purification or subsequent handling and storage. A region may also be added to the polypeptide to facilitate purification. Such regions may be removed prior to final preparation of the polypeptide. The addition of peptide moieties to polypeptides to engender secretion or excretion, to improve stability and to facilitate purification, among others, are familiar and routine techniques in the art.

Suitable prokaryotic hosts for propagation, maintenance or expression of polynucleotides and polypeptides in accordance with the invention include *Escherichia coli*, *Bacillus subtilis* and *Salmonella typhimurium*. Various species of *Pseudomonas*, *Streptomyces*, and *Staphylococcus* are also suitable hosts in this regard. Moreover, many other hosts also known to those of skill may be employed in this regard.

As a representative but non-limiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids comprising genetic elements of the well known cloning vector pB22 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM1 (Promega Biotec, ison, WI, USA). In these vectors, the pB22 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed.

Following transformation of a suitable host strain, the host strain is grown to an appropriate cell density. Where the selected promoter is inducible, it is induced by appropriate means (e.g., temperature shift or exposure to chemical inducer) and cells are cultured for an additional period. Cells typically then are harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents. Such methods are well known to those skilled in the art.

Various mammalian cell culture systems can be employed for expression, as well. Examples of mammalian expression systems include, without limitation, the C127, 3T3, CHO, HeLa, human kidney 293 and BHK cell lines, and the COS-7 line of monkey kidney fibroblasts, described by Gluzman et al., Cell, 1981, 23:175. Mammalian expression vectors will comprise an origin of replication, a suitable promoter and enhancer, and any necessary ribosome binding sites, polyadenylation sites, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking non-transcribed sequences that are necessary for expression. In certain preferred embodiments, DNA sequences derived from the SV40 splice sites and the SV40 polyadenylation sites are used for required non-transcribed genetic elements.

The polypeptide or fusion protein of this invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography ("HPLC") is employed for purification. Well known techniques for refolding proteins may be employed to regenerate active conformation when the polypeptide is denatured during isolation and or purification.

V. Uses of the Polypeptides and Polynucleotides of this Invention

The polynucleotides and polypeptides of the present invention may be used in accordance with the present invention for a variety of applications, particularly in the

development of vaccines and as diagnostics. As immunogenic compositions, for vaccine use or for the development of antibodies, the polypeptides of this invention are preferably used as fusion proteins. The polynucleotides and polypeptides may occur in a composition, such as a media, formulations, solutions for introduction of polynucleotides or polypeptides, for example, into cells, compositions or solutions for chemical or enzymatic reactions, for instance, which are not naturally occurring compositions, and, therein remain isolated polynucleotides or polypeptides within the meaning of that term as it is employed herein.

As diagnostic compositions, the polypeptides are preferably used as individual peptides, or peptides coupled to a polylysine core (so-called multiple antigenic peptides), and/or labeled peptides.

These polynucleotides and polypeptides may also be employed in the development of binding molecules that interfere with the binding of the bacterial adhesin to its ligand, and in the use thereof as pharmaceutical agents. Additional applications relate to diagnosis and to treatment of disorders of cells, tissues and organisms. These aspects of the invention are illustrated further by the following discussion.

A. *Polynucleotide assays*

This invention is also related to the use of the polynucleotides described above to detect complementary polynucleotides for use, for example, as a diagnostic reagent. Detection of one of the conserved sequences of a bacterial adhesin identified above provides a diagnostic tool that can add to or define diagnosis of an infection with a proteobacterial species.

Nucleic acids for diagnosis may be obtained from a patient's cells, such as from blood, urine, saliva, tissue biopsy or autopsy material. The genomic DNA may be used directly for detection or may be amplified enzymatically by using polymerase chain reaction (PCR) [Saiki et al., Nature, 324:163-166 (1986)] prior to analysis. RNA or cDNA may also be used in similar fashion. As an example, PCR primers complementary to the nucleic acid encoding a polypeptide of this invention can be used to identify and analyze expression of a bacterial adhesin.

Sequence differences between a reference adhesin gene and genes having mutations may also be revealed by direct DNA sequencing. In addition, cloned

DNA segments may be employed as probes to detect specific DNA segments. The sensitivity of such methods can be greatly enhanced by appropriate use of PCR or other amplification methods. For example, a sequencing primer is used with double-stranded PCR product or a single-stranded template molecule generated by a modified PCR. The sequence determination is performed by conventional procedures with radiolabeled nucleotide or by automatic sequencing procedures with fluorescent-tags.

Sequence changes at specific locations may also be revealed by nuclease protection assays, such as RNase and S1 protection or the chemical cleavage method [e.g., Cotton et al., Proc. Natl. Acad. Sci., USA, 85:4397-4401 (1985)].

Thus, the detection of a specific bacterial adhesin DNA sequence may be achieved by methods such as hybridization, RNase protection, chemical cleavage, direct DNA sequencing or the use of restriction enzymes, (e.g., restriction fragment length polymorphisms ("RFLP"), PCR, RT-PCR, Northern blotting and Southern blotting, and *in situ* analysis.

B. Polypeptide assays

The present invention also relates to diagnostic assays for detecting, qualitatively or quantitatively, the presence of proteobacterial adhesin protein in infected cells and tissues. Assay techniques that can be used to determine levels of a protein, such as conserved polypeptide of the present invention, in a sample derived from a host are well-known to those of skill in the art. Such assay methods include radioimmunoassays, competitive-binding assays, Western Blot analysis and ELISA assays. Among these, ELISAs are frequently preferred. An ELISA assay initially comprises preparing an antibody specific to a polypeptide of this invention or , preferably a monoclonal antibody. In addition a reporter antibody generally is prepared which binds to the monoclonal antibody. The reporter antibody is attached to a detectable reagent such as a radioactive, fluorescent or enzymatic reagent, e.g., horseradish peroxidase enzyme.

To carry out an ELISA, a sample is removed from a host and incubated on a solid support, e.g. a polystyrene dish, that binds the proteins in the sample. Any free protein binding sites on the dish are then covered by incubating with a non-specific protein such as bovine serum albumin. The monoclonal antibody is then incubated in the dish during which time the monoclonal antibodies attach to any

proteobacterial adhesin sequences attached to the polystyrene dish. Unbound monoclonal antibody is washed out with buffer. The reporter antibody linked to horseradish peroxidase is placed in the dish resulting in binding of the reporter antibody to any monoclonal antibody bound to the adhesin sequence. Unattached reporter antibody is then washed out. Reagents for peroxidase activity, including a colorimetric substrate, are then added to the dish. Immobilized peroxidase, linked to or through the primary and secondary antibodies, produces a colored reaction product. The amount of color developed in a given time period indicates the amount of the adhesin present in the sample. Quantitative results typically are obtained by reference to a standard curve.

A competition assay may also be employed to determine levels of the polypeptide of the present invention in a sample derived from the infected hosts. Such an assay comprises isolating cells which express the polypeptide of the present invention. A test sample containing the polypeptides of the present invention which have been labeled, are then added to the purified cells and then incubated for a set period of time. Also added to the reaction mixture is a sample derived from a host which is suspected of containing the polypeptide of the present invention. The reaction mixtures are then passed through a filter which is rapidly washed and the bound radioactivity is then measured to determine the amount of competition for the polypeptides and therefore the amount of the polypeptides of the present invention in the sample.

Another competition assay may involve antibodies specific to a polypeptide of the invention, which are attached to a solid support and labeled or and a sample derived from the host are passed over the solid support. The amount of detected label attached to the solid support can be correlated to a quantity of a bacterial adhesin sequence in the sample.

C. Development of Immunogenic Binding Molecules

"Binding molecules" (or otherwise called "interaction molecules" or "receptor component factors") refer to molecules, including ligands, that specifically bind to or interact with polypeptides of the present invention. Such binding molecules are a part of the present invention. Binding molecules may also be non-naturally occurring, such as antibodies and antibody-derived reagents that bind specifically to polypeptides of the invention.

1. Antibodies

The polypeptides of this invention, their fragments or other derivatives, or analogs thereof, or cells expressing them can also be used as immunogens to produce antibodies thereto. These antibodies can be, for example, polyclonal or monoclonal antibodies. The present invention also includes chimeric, single chain, and humanized antibodies, as well as Fab fragments, or the product of an Fab expression library. Various procedures known in the art may be used for the production of such antibodies and fragments.

Antibodies generated against the polypeptides corresponding to a sequence of the present invention can be obtained by direct injection of the polypeptides into an animal or by administering the polypeptides to an animal, preferably a nonhuman. The antibody so obtained will then bind the polypeptide itself. In this manner, even a sequence encoding only a fragment of the polypeptide can be used to generate antibodies binding the whole native polypeptide. Such antibodies can then be used to isolate the polypeptide from tissue expressing that polypeptide.

For preparation of monoclonal antibodies, any technique which provides antibodies produced by continuous cell line cultures can be used. Examples include the hybridoma technique [G. Kohler and C. Milstein, Nature, 256:495-497 (1975)], the trioma technique, the human B-cell hybridoma technique [Kozbor et al., Immunology Today, 4:72 (1983)], and the EBV-hybridoma technique [Cole et al., MONOCLONAL ANTIBODIES AND CANCER THERAPY, pg. 77-96, Alan R. Liss, Inc., (1985)].

Techniques described for the production of single chain antibodies [U.S. Patent No. 4,946,778] can also be adapted to produce single chain antibodies to immunogenic polypeptide products of this invention. Also, transgenic rabbits, or other organisms including other mammals, may be used to express humanized antibodies to immunogenic polypeptide products of this invention.

The above-described antibodies may be employed to isolate or to identify clones expressing the polypeptide or purify the polypeptide of the present invention by attachment of the antibody to a solid support for isolation and/or purification by affinity chromatography. Antibodies, including monoclonal antibodies, against the

polypeptides of this invention may also be employed to treat or prohibit proteobacterial infections.

2. *Binding Molecules and Assays*

A polypeptide of this invention can be used to isolate proteins which interact with it, or to identify its ligand when expressed on an infected host cell; and this interaction can be a target for interference. Inhibitors of protein-protein interactions between the conserved Proteobacterial polypeptide of the and other factors could lead to the development of pharmaceutical agents for the treatment of infection by such bacterial species.

Thus, this invention also provides a method for identification of binding molecules to the conserved polypeptides of this invention. Polynucleotide sequences encoding proteins for binding molecules to the polypeptides of this invention can be identified by numerous methods known to those of skill in the art, for example, ligand panning and FACS sorting. Such methods are described in many laboratory manuals such as, for instance, Coligan et al., CURRENT PROTOCOLS IN IMMUNOLOGY 1, Chapter 5 (1991).

For example, the yeast two-hybrid system provides methods for detecting the interaction between a first test protein and a second test protein, *in vivo*, using reconstitution of the activity of a transcriptional activator. The method is disclosed in U.S. Patent No. 5,283,173; reagents are available from Clontech and Stratagene. Briefly, a polynucleotide sequence encoding a conserved polypeptide of this invention is fused to a Gal4 transcription factor DNA binding domain and expressed in yeast cells. cDNA library members obtained from cells of interest are fused to a transactivation domain of Gal4. cDNA clones which express proteins which can interact with the conserved polypeptide sequence will lead to reconstitution of Gal4 activity and transactivation of expression of a reporter gene such as Gal1-lacZ.

An alternative method involves screening of lambda gt11 or lambda ZAP (Stratagene) or equivalent cDNA expression libraries with recombinant polypeptides of this invention. Recombinant polypeptides of this invention are fused to small peptide tags such as FLAG, HSV or GST. The peptide tags can possess convenient phosphorylation sites for a kinase such as heart muscle creatine kinase or they can be

biotinylated. Recombinant polypeptides of this invention can be phosphorylated with ^{32}[P] or used unlabeled and detected with streptavidin or antibodies against the tags. Lambda gt11 cDNA expression libraries are made from cells of interest and are incubated with the recombinant polypeptide of this invention, washed and cDNA clones which interact with the polypeptide are isolated. Such methods are routinely used by skilled artisans. See, e.g., Sambrook et al. cited above.

Another method for obtaining molecules that bind the polypeptides of the present invention involves the screening of a mammalian expression library. In this method, cDNAs are cloned into a vector between a mammalian promoter and polyadenylation site and transiently transfected in COS or 293 cells. Forty-eight hours later, the binding protein is detected by incubation of fixed and washed cells with labeled polypeptide. In a preferred embodiment, the polypeptide of this invention is iodinated, and any bound polypeptide is detected by autoradiography. See Sims *et al.*, Science, 1988, 241:585-589 and McMahan *et al.*, EMBO J., 1991, 10:2821-2832. In this manner, pools of cDNAs containing the cDNA encoding the binding protein of interest can be selected and the cDNA of interest can be isolated by further subdivision of each pool followed by cycles of transient transfection, binding and autoradiography. Alternatively, the cDNA of interest can be isolated by transfecting the entire cDNA library into mammalian cells and panning the cells on a dish containing a polypeptide of this invention bound to the plate. Cells which attach after washing are lysed and the plasmid DNA isolated, amplified in bacteria, and the cycle of transfection and panning repeated until a single cDNA clone is obtained. See Seed *et al*, Proc. Natl. Acad. Sci. USA, 1987, 84:3365 and Aruffo *et al.*, EMBO J., 1987, 6:3313. If the binding protein is secreted, its cDNA can be obtained by a similar pooling strategy once a binding or neutralizing assay has been established for assaying supernatants from transiently transfected cells. General methods for screening supernatants are disclosed in Wong *et al.*, Science, 1985, 228:810-815.

Another method of identifying a binding molecule involves isolation of proteins interacting with a polypeptides of this invention directly from cells infected with a proteobacterial species. Fusion proteins of a polypeptide of this invention with GST or small peptide tags are made and immobilized on beads. Biosynthetically labeled polypeptide of this invention unlabeled protein extracts from the cells of interest

are prepared, incubated with the beads and washed with buffer. Proteins interacting with the polypeptide of the invention are eluted specifically from the beads and analyzed by SDS-PAGE. Binding partner primary amino acid sequence data are obtained by microsequencing. Optionally, the cells can be treated with agents that induce a functional response such as tyrosine phosphorylation of cellular proteins. An example of such an agent would be a growth factor or cytokine such as interleukin-2.

Another method for identifying binding molecules is immunoaffinity purification. A recombinant polypeptide of this invention is incubated with a labeled polypeptide of this invention, unlabeled cell extracts and immunoprecipitated with antibodies to the polypeptide of this invention. The immunoprecipitate is recovered with protein A-Sepharose and analyzed by SDS-PAGE. Unlabelled proteins are labeled by biotinylation and detected on SDS gels with streptavidin. Binding partner proteins are analyzed by microsequencing. Further, standard biochemical purification steps known to those skilled in the art may be used prior to microsequencing.

Yet another alternative method involves screening of peptide libraries for binding partners. Recombinant tagged or labeled polypeptides of this invention are used to select peptides from a peptide or phosphopeptide library which interact with the polypeptides of the invention. Sequencing of the peptides leads to identification of consensus peptide sequences which might be found in interacting proteins.

Another method for identifying compounds which antagonize the binding of a bacterial adhesin to its ligand comprise the steps of providing a sample of the ligand or a cell which expresses the ligand immobilized on a support; contacting the sample with a known amount of a polypeptide of this invention and a known amount of a test compound; washing unbound materials from the sample; contacting the sample with a labeled reagent which binds to the polypeptide; washing unbound reagent from the sample; measuring the amount of signal generated by the label. The amount of signal generated is inversely proportional to the ability of the test compound to disrupt or inhibit binding between the polypeptide and the ligand; and identifying those test compounds as antagonists which are associated with a low signal.

The binding partners or antagonists identified by any of these methods or other methods, which would be known to those of ordinary skill in the art, as well as those putative binding partners discussed above, can be used in the assay method of the invention. Assaying for the presence of the native conserved polypeptide/binding partner complex is accomplished by, for example, the yeast two-hybrid system, ELISA or immunoassays using antibodies specific for the complex. In the presence of test substances which interrupt or inhibit formation of a conserved Proteobacterial sequence/binding partner interaction, a decreased amount of complex will be determined relative to a control lacking the test substance.

Assays for free polypeptide or binding partner are accomplished by, for example, ELISA or immunoassay using specific antibodies or by incubation of radiolabeled or with cells or cell membranes followed by centrifugation or filter separation steps. In the presence of test substances which interrupt or inhibit formation of the interaction between a Proteobacterial conserved sequence and a binding partner, an increased amount of free conserved polypeptide or free binding partner will be determined relative to a control lacking the test substance.

Polypeptides of the invention also can be used to assess or binding capacity of the polypeptides of this invention or their binding molecules in cells or in cell-free preparations. Other methods for detecting ligands (agonists or antagonists) for the polypeptides of this invention include the yeast based technology as described in U. S. Patent No. 5,482,835. Examples of potential ligands include antibodies or, in some cases, oligonucleotides which bind to the polypeptide. Potential antagonists also include proteins which are closely related to a ligand of the polypeptide of this invention, i.e., a fragment of the ligand. A potential antagonist also includes an antisense construct prepared through the use of antisense technology. Antisense technology can be used to control gene expression through triple-helix formation or antisense DNA or RNA, both methods of which are based on binding of a polynucleotide to DNA or RNA. For example, the 5' coding portion of the polynucleotide sequence, which encodes for the polypeptides of the present invention, is used to design an antisense RNA oligonucleotide of from about 10 to 24 base pairs in length [see, Lee et al., Nucl. Acids Res., 6:3073 (1979); Cooney et al., Science, 241:456 (1988); Dervan et al., Science, 251:1360 (1991);

Okano, J. Neurochem., 56:560 (1991); and OLIGODEOXYNUCLEOTIDES AS ANTISENSE INHIBITORS OF GENE EXPRESSION, CRC Press, Boca Raton, FL (1988)]. The oligonucleotides described above can also be delivered to cells such that the antisense RNA or DNA is expressed *in vivo* to inhibit the binding of the conserved sequence to its ligand in infected cells.

Another potential antagonist is a small molecule which binds to a conserved Proteobacterial sequence of this invention, making it inaccessible to ligands such that normal biological activity is prevented. Examples of small molecules include, but are not limited to, small peptides or peptide-like molecules. The small molecules may also bind the interaction protein to the sequence. One exemplary method for generating a small molecule which antagonizes the binding between a proteobacterial adhesin and its ligand involves analyzing an antibody to a polypeptide as described above in a computer modelling program.

Potential antagonists also include fragments of the polypeptides of the invention, which bind to the ligand and prevent the ligand from interacting with the conserved Proteobacterial sequence in infected cells. It is desirous to find compounds and drugs which can inhibit the function of the conserved Proteobacteral sequence. In general, agonists or antagonists for the polypeptides of this invention are employed for diagnostic, therapeutic and prophylactic purposes for the diagnosis or treatment of infection by the Proteobacterial species identified herein, among others.

For example, a process for diagnosing a bacterial infection comprises contacting a biological sample from a possibly infected subject with a labeled antibody which binds to the conserved polypeptide of the sequences of the invention set forth herein; and measuring the signal generated by the label with a suitable assay, wherein detection of the signal indicates the presence of an adhesin molecule from the bacteria. Thus, a diagnostic reagent of this invention comprises a composition capable of binding to one of the polypeptides of the sequences of the invention set forth herein, the composition associated with a detectable label.

For use in diagnostic assays, the polypeptides, fusion proteins and/or other reagents of the invention identified above are associated with conventional labels which are capable, alone or in concert with other compositions or

compounds. of providing a detectable signal. The labels may be interactive to produce a detectable signal. Most desirably, the label is detectable visually, e.g. colorimetrically. A variety of enzyme systems have been described in the art which will operate to reveal a colorimetric signal in an assay. As one example, glucose oxidase (which uses glucose as a substrate) releases peroxide as a product. Peroxidase, which reacts with peroxide and a hydrogen donor such as tetramethyl benzidine (TMB) produces an oxidized TMB that is seen as a blue color. Other examples include horseradish peroxidase (HRP) or alkaline phosphatase (AP), and hexokinase in conjunction with glucose-6-phosphate dehydrogenase which reacts with ATP, glucose, and NAD+ to yield, among other products, NADH that is detected as increased absorbance at 340 nm wavelength. Other label systems that may be utilized in the methods of this invention are detectable by other means, e.g., colored latex microparticles [Bangs Laboratories, Indiana] in which a dye is embedded may be used in place of enzymes to form conjugates with the antibodies and provide a visual signal indicative of the presence of the resulting complex in applicable assays. Still other labels include fluorescent compounds, radioactive compounds or elements. Detectable labels for attachment to polypeptides, proteins, and antibodies useful in diagnostic assays of this invention may be easily selected from among numerous compositions known and readily available to one skilled in the art of diagnostic assays. The methods and antibodies of this invention are not limited by the particular detectable label or label system employed.

It should be understood by one of skill in the art that any number of conventional protein assay formats, particularly immunoassay formats, or nucleic acid assay formats, may be designed to utilize the isolated polypeptides, fusion proteins, antibodies, binding moleucles or their nucleic acid sequences or anti-sense sequences of this invention for the detection of *Proteobacterial* infection in animals and humans. This invention is thus not limited by the selection of the particular assay format, and is believed to encompass assay formats which are known to those of skill in the art.

For convenience, reagents for ELISA or other assays according to this invention may be provided in the form of kits. Such kits are useful for diagnosing infection with *Proteobacterial* species in a human or an animal sample. Such a diagnostic kit contains an antigen of this invention and/or at least one polypeptide, fusion

protein, or antibody capable of binding a Proteobacterial sequence identified by this invention, or the nucleic acid sequences encoding them, or their anti-sense sequences. Alternatively, such kits may contain a simple mixture of such antigens or sequences, or means for preparing a simple mixture.

These kits can include microtiter plates to which the Proteobacterial antigen proteins or antibodies or nucleic acid sequences of the invention have been pre-adsorbed, various diluents and buffers, labeled conjugates for the detection of specifically bound antigens or antibodies, or nucleic acids and other signal-generating reagents, such as enzyme substrates, cofactors and chromogens. Other components of these kits can easily be determined by one of skill in the art. Such components may include polyclonal or monoclonal capture antibodies, antigen of this invention, or a cocktail of two or more of the antibodies, purified or semi-purified extracts of these antigens as standards, MAb detector antibodies, an anti-mouse or anti-human antibody with indicator molecule conjugated thereto, an ELISA plate prepared for absorption, indicator charts for colorimetric comparisons, disposable gloves, decontamination instructions, applicator sticks or containers, and a sample preparator cup. Such kits provide a convenient, efficient way for a clinical laboratory to diagnose *Proteobacterial* infection.

D. Vaccine Uses

Thus in one aspect, this invention provides an immunogenic composition useful as a vaccine to prevent infection by a proteobacterial species comprising in a pharmaceutically acceptable carrier, at least one component described above, such as, a polypeptide derived from a conserved sequence of a Proteobacterial species, e.g., the sequences of the invention set forth herein; an amino acid sequence at least 50% identical to the polypeptide sequence as determined by a sequence comparison algorithm, which sequence binds the ligand of the polypeptide sequence. Preferably these polypeptides of the present invention are employed as fusion proteins comprising the Proteobacterial polypeptide described above fused in frame to a second protein for vaccine use. Such fusion proteins are described in detail above. The polypeptides of this invention are administered, desirably as fusion proteins, to develop in a mammalian subject *in vivo*, antibodies to the conserved polypeptides sequences of the infecting Proteobacterial

species. In this manner. the polypeptides of this invention are useful as vaccine components. These polypeptides are administered in an amount effective to induce a humoral or cellular immune response against the invading bacteria in a manner so as to inhibit the infection by blocking binding of ligands to the conversed polypeptide of the invention.

Additionally, a small molecule which binds the ligand of the polypeptide sequence; an antibody which binds the polypeptide sequence; or an anti-idiotype antibody of the aforementioned antibody may be employed in vaccine compositions or in therapeutic compositions. In any of these compositions, an optional adjuvant may be included of which many types are available for selection by one of skill in the pharmaceutical arts.

In one embodiment, this invention additionally provides a method of treating an infection by a Proteobacteral species bacteria which comprises administering to a subject an inhibitor compound (antagonist) as herein above-described along with a pharmaceutically acceptable carrier in an amount effective to inhibit the spread of infection by blocking binding of ligands to the conversed polypeptide of the invention.

In another embodiment of this invention, a binding molecule, preferably an antibody, developed as described above, may be employed as a passive vaccine to prevent infection by a Proteobacteral species. According to this aspect, a mammalian subject is administered an antibody or cocktail of antibodies in a suitable pharmaceutical carrier and with optional adjuvants to the polypeptides of this invention prior to infection to provide passive prophylaxsis whenever exposure to such bacterial species is contemplated. These antibodies are administered in an amount effective to inhibit the spread of infection by blocking binding of ligands to the conversed polypeptide of the invention.

I. Compositions

These polypeptides of the invention, and compounds which bind or inhibit interaction between the Proteobacterial conserved sequences and their ligands, may be employed in combination with a suitable pharmaceutical, physiologically acceptable carrier. For example, one such vaccine composition may be formulated to contain a carrier or diluent and one or more of the polypeptide/fusion protein or multimeric

proteins of the invention. Suitable pharmaceutically acceptable carriers facilitate administration of the proteins but are physiologically inert and/or nonharmful. Carriers may be selected by one of skill in the art. Such carriers include but are not limited to, sterile saline, phosphate, buffered saline, dextrose, sterilized water, glycerol, ethanol, lactose, sucrose, calcium phosphate, gelatin, dextran, agar, pectin, peanut oil, olive oil, sesame oil, and water and combinations thereof. Additionally, the carrier or diluent may include a time delay material, such as glycerol monostearate or glycerol distearate alone or with a wax. In addition, slow release polymer formulations can be used. The formulation should suit the mode of administration. Selection of an appropriate carrier in accordance with the mode of administration is routinely performed by those skilled in the art.

Optionally, the vaccine composition may further contain adjuvants, preservatives, chemical stabilizers, or other antigenic proteins. Typically, stabilizers, adjuvants, and preservatives are optimized to determine the best formulation for efficacy in the target human or animal. Suitable exemplary preservatives include chlorobutanol, potassium sorbate, sorbic acid, sulfur dioxide, propyl gallade, the parabens, ethyl vanillin, glycerin, phenol, and parachlorophenol. Suitable stabilizing ingredients which may be used include, for example, casamino acids, sucrose, gelatin, phenol red, N-Z amine, monopotassium diphosphate, lactose, lactalbumin hydrolysate, and dried milk.

One or more of the above described vaccine components may be admixed or adsorbed with a conventional adjuvant. The adjuvant is used to attract leukocytes or enhance an immune response. Such adjuvants include, among others, Ribi, mineral oil and water, aluminum hydroxide, Amphigen, Avridine, L121/squalene, D-lactide-polylactide/glycoside, pluronic plyois, muramyl dipeptide, killed *Bordetella*, and saponins, such as Quil A. Other vaccinal antigens originating from other bacterial species may also be included in these compositions.

Polypeptides and other compounds of the present invention which inhibit the interaction between the conserved sequence of the Proteobacterial species and its ligand may be employed alone or in conjunction with other compounds, such as therapeutic compounds. In addition to the polypeptides of the invention, other agents useful in treating a Proteobacterial infection, e.g., antibiotics or immunostimulatory agents and cytokine regulation elements, are expected to be useful in preventing, reducing

or eliminating disease symptoms. Such agents may operate in concert with the compositions of this invention.

The invention further relates to pharmaceutical packs and kits comprising one or more containers filled with one or more of the ingredients of the aforementioned compositions of the invention.

2. *Vaccine/Therapeutic Administration*

Infection by the Proteobacterial species may be partially or completely ameliorated by the systemic clinical administration of the polypeptides/antibodies of this invention, or by administration as a vaccine and again as multiple boosters, as required. This administration can be through the administration of peptides agonists or antagonists synthesized from recombinant constructs of polynucleotides encoding the polypeptide of this invention or from peptide chemical synthesis [see, e.g., Woo *et al.*, Protein Engineering, 3:29-37 (1989)]. The pharmaceutical compositions may be administered in any effective, convenient manner including, for instance, administration by topical, oral, anal, vaginal, intravenous, intraperitoneal, intramuscular, subcutaneous, intranasal or intradermal routes, among others.

According to the method of the invention, a human or an animal may be vaccinated against Proteobacterial infection by administering an effective amount of such a composition. An "effective amount" is defined as an amount of antigen that is effective in a route of administration to provide a vaccinal benefit, i.e., protective immunity. Such an amount may be between about 1 ng to 1000 mg protein, and more preferably, 0.05 μg to 1 mg per mL of protein; or 0.05 to about 1000 μg/mL of a polypeptide or fusion protein of the invention. A suitable dosage may be about 1.0-5.0 mL of a vaccine composition. Suitable dosage adjustments and the need for any boosters may be made by the attending physician or veterinarian depending upon the age, sex, weight and general health of the human or animal patient, as well as the level of immune response desired. The vaccine may be administered by any suitable route. Preferably, such a composition is administered parenterally, preferably intramuscularly or subcutaneously. However, it may also be formulated to be administered by any other suitable route, including orally or topically. Routes of administration may be combined, if desired, or

adjusted.

Further, the vaccine may be a DNA vaccine, which includes a nucleotide sequence encoding one or more of the polypeptides or fusion proteins of this inveniton, optionally under the control of regulatory sequences. Thus, the antigen-encoding DNA may be carried in a vector, e.g., a viral vector. Generally, a suitable vector-based treatment contains between 1×10^{-3} pfu to 1×10^{12} pfu per dose. However, the dose, timing and mode of administration of these compositions may be determined by one of skill in the art. Such factors as the age, and physical condition of the vaccinate may be taken into account in determining the dose, timing and mode of administration of the immunogenic or vaccine composition of the invention.

While the above dosage ranges are guidelines only, in general, the pharmaceutical compositions generally are administered in an amount effective for treatment or prophylaxis of infection by a bacterial species described herein. The amount employed of the subject polypeptide or binding compound will vary with the manner of administration, the employment of other active compounds, and the like. Another conventional general range is about 1 μg to 100 μg. The amount of compound employed will be determined empirically, based on the response of cells *in vitro* and response of experimental animals to the subject polypeptides or formulations containing the subject polypeptides. In general, the compositions are administered in an amount of at least about 10 μg/kg body weight. In most cases they will be administered in an amount not in excess of about 8 mg/kg body weight per day. Preferably, in most cases, the administered dose is from about 10 μg/kg to about 1 mg/kg body weight, daily. It will be appreciated that optimum dosage will be determined by standard methods for each treatment modality and indication, taking into account the indication, its severity, route of administration, complicating conditions and the like.

EXAMPLES

The present invention is further described by the following examples, which are provided solely to illustrate the invention by reference to specific embodiments. These exemplifications do not limit or circumscribe the scope of the disclosed invention.

Certain terms used herein are explained in the foregoing glossary. All examples are carried out using standard techniques, which are well known and routine to those of skill in the art, except where otherwise described in detail. Routine molecular biology techniques of the following examples can be carried out as described in standard laboratory manuals, such as Sambrook et al. All parts or amounts set out in the following examples are by weight, unless otherwise specified.

Unless otherwise stated, size separation of fragments in the examples below is carried out using standard techniques of agarose and polyacrylamide gel electrophoresis ("PAGE") as described in Sambrook et al and numerous other references, such as D. Goeddel *et al.*, Nucleic Acids Res., 8: 4057 (1980) (i.e., using 8 percent polyacrylamide gel). Unless described otherwise, ligations are accomplished using standard buffers, incubation temperatures and times, e.g., approximately 10 units of T4 DNA ligase ("ligase") per 0.5 μg of approximately equimolar amounts of the DNA fragments to be ligated.

EXAMPLE 1 - SYNTHESIS OF A POLYPEPTIDE OF THIS INVENTION AND METHODS FOR OBTAINING AN ANTIBODY THERETO

A polypeptide corresponding to the sequences of the invention set forth herein is synthesized as described below. The amino acid sequence of this immunogen is synthesized by solid phase methodology on polypropylene pegs according to the methods of H. M. Geysen *et al.*, J. Immunol. Meth., 102:259 (1987), with an N-terminal cysteinyl being incorporated to facilitate coupling to a carrier protein. The N-terminus is left as a free amine and the C-terminus was amidated in the immunizing polypeptides. Immunizing polypeptides are generally purified to greater than 95% purity by reverse phase HPLC, and purity is further confirmed by mass spectometry (MS).

Immunizing polypeptides are covalently coupled to diphtheria toxoid (DT) carrier protein via the cysteinyl side chain by the method of A. C. J. Lee *et al.*, Molec. Immunol., 17:749 (1980), using a ratio of 6-8 moles peptide per mole of diphtheria toxoid.

The polypeptide conjugates are taken up in purified water and emulsified 1:1 with complete Freund's adjuvant (CFA) or incomplete Freund's adjuvant (IFA) [ANTIBODIES - A LABORATORY MANUAL, Eds. E. Harlow and P. Lane, Cold Spring

Harbor Laboratory (1998)]. Total volume per immunized rabbit is 1 ml. and this contains 100 µg of peptide coupled to DT.

Five rabbits are used for the immunizing peptide, with the initial intramuscular (IM) injection with conjugate in CFA and a subsequent IM boost at 2 weeks with conjugate in IFA. A pre-bleed is drawn before the first injection and larger bleeds are taken 3 and 5 weeks after the booster injection.

These assays are performed as described by H.M. Geysen *et al.*, Proc. Natl. Acad. Sci. USA, 81:3998 (1983). Briefly, using Nunc Immuno Maxisorbϑ 96 well plates, biotinylated polypeptides are bound to streptavidin coated plates and, with washing with phosphate buffered saline (PBS) between steps, successive incubations are performed with antiserum dilutions and horseradish peroxidase conjugated anti-rabbit immunoglobulin to detect bound antibody. Plates are developed with ABTS, with an O.D. reading at 405 nm. Absorbance greater than O.D. 1.0 is taken as positive and titers are determined from doubling dilutions of each antiserum. The geometric mean titer (GMT) is calculated for each antiserum pair for a given immunizing polypeptide.

EXAMPLE 2: RECOMBINANT EXPRESSION OF A POLYPEPTIDE OF THIS INVENTION

The DNA sequence encoding a polypeptide or fusion protein of this invention is cut from a Bluescript plasmid using the restriction enzyme sites corresponding to the restriction enzyme sites on the bacterial expression vector pBluescript SK (+/-) phagemid (Stratagene, Inc.). pBluescript SK (+/-) phagemid encodes antibiotic resistance (Ampr), a bacterial origin of replication (ori), an ß-galactosidase promoter operator, and other regulatory sequences [GENBANK 52325].

Plasmid GEX-t1 [Pharmacea, Uppsala, Sweden] is then digested with EcoRI and XhoI and the polynucleotide sequence encoding a polypeptide sequence of this invention is ligated into the digested plasmid. The polypeptide-encoding sequence is inserted in frame with the sequence encoding for the glutathione S transferase gene in this commercially available plasmid. This plasmid is designed to generate fusion of the inserted polypeptide-encoding sequence, with GST. The ligation mixture is then used to transform E. coli strain SOLR (Stratagene) by conventional techniques. The fusion protein

polypeptide-GST is purified using GST Sephadex (Pharmacea) according to manufacturer's instructions.

EXAMPLE 3 - ASSAY PROCEDURES

Enzyme assay procedures for identifying agonists and antagonists of a polypeptide of this invention include assays which use the FLASHPLATE system (DuPont), as follows. The approach taken for the measurement of binding activity: one uses the coating of an antibody to the polypeptide of this invention, coated onto plates directly. FlashPlate is coated with 100 µL per well of an antibody against a polypeptide the sequences of the invention set forth herein at a concentration of 5 µg/mL in PBS. After an overnight incubation at room temperature, the plate was washed twice with PBS and then blocked with 1% BSA/PBS for at least 2 hours at room temperature. The plate was air dried and stored at 4EC until use. Plates are viable for 2-3 weeks when stored at 4EC.

The phosphorylation reaction was performed in the plate using a total volume of 60 µL per well containing 33 mM Tris-HCl (pH 7.4), 17 mM $MgCl_2$, 33 µM ATP, 0.7 mM DTT, 0.25 µCi of [γ^{33}P]-ATP (DuPont NEG-302H), 20 µg of and varying amounts of purified polypeptide or fusion protein. The plate is incubated overnight at 30EC. Following aspiration of the solution, the wells are rinsed 1x with 250 µL per well of 10 mM sodium pyrophosphate/PBS which reduces non-specific binding. The plate is counted on a Packard TopCount.

The protein immobilized directly onto FlashPlate serves as a functional substrate for the conserved Proteobacterial sequence. The reaction only require a single pyrophosphate rinse to remove unreacted [γ^{33}P]-ATP and cell lysate from the wells. Background counts in wells containing no Proteobacterial sequence has a characteristic count and a signal to noise ratio. This ratio increases as the amount of interaction between the antibody on the plate and the polypeptide in the cell lysate increases. Immobilized substrate at 750 ng/well can be phosphorylated in a dose dependent fashion, thus allowing quantitation of binding activity.

Coating the plate with an antibody against the polypeptide is also efficient in enabling the bound substrate to be phosphorylated by the Proteobacteral

sequences in the lysate. The reaction is dose dependent with respect to the amount of Proteobacterial sequences added.

Various options are available for formatting an enzyme assay. Such assays enable one to insert into the system an unknown compound, which can inhibit the binding reaction by interacting with the polypeptide of the invention or with its ligand expressed by an infected cell. The choice of format depends upon the sensitivity required and the purpose of the assay. Regardless of format, such enzyme assays are advantageous both for automation and for high throughput screening.

EXAMPLE 4 - ANIMAL STUDY

A study is conducted in ten rabbits to determine if the presence of antibodies to a conserved sequence of this invention, induced by a synthetic peptide of this invention (*e.g.*, the sequences of the invention set forth herein) prior to infection with a Proteobacterial species, e.g., *Yersinia pestis*, attenuates or reduces infection.

A. Immunization of rabbits

The rabbits are randomized into two groups. Each rabbit of group 1 (control group) is immunized with 0.4 mg diphtheria toxoid (Commonwealth Serum Laboratories, Victoria, Australia) with 0.25 mg threonyl muramyl dipeptide (T-MDP) in 0.5 ml water, this being emulsified with 0.5 ml MF75 adjuvant (Chiron Corp, Emeryville CA). Each rabbit of group 2 (test group) is immunized with 0.1 mg of the synthetic polypeptide of the invention (*e.g.*, the sequences of the invention set forth herein) in which R is H and R^2 is an amide coupled to 0.4 mg diphtheria toxoid [A. C. Lee *et al.*, Mol. Immunol., 17:749 (1980)]. The conjugate is dissolved in 0.5 ml water containing 0.25 mg T-MDP and emulsified with 0.5 ml MF75 adjuvant.

Each rabbit is immunized at day 0 and day 28 (week 4) with two 0.5 ml intramuscular injections at two distinct sites. At day 42 (week 6), 2 weeks after the booster injection, serums are drawn and tested by ELISA for binding to a sequence of the invention set forth herein, as described above in Example 1. This assay indicates the background titers of the control rabbits and the titers of the test group.

At day 49 (week 7) after initial immunization, all rabbits are given 50 animal infectious doses$_{50\%}$ (50 AID_{50}) or 200 tissue culture infectious doses$_{50\%}$ (200

$TCID_{50}$) of *Yersinia pestis*. Plasma is drawn in EDTA at weeks 2, 4 and 8, and copies of viral RNA per ml of plasma were measured by RT-PCR.

Yersinia pestis in control animals causes a characteristic infection. Rabbits immunized with a synthetic polypeptide according to this invention that induce antibodies to the conserved sequence of the *Yersinia pestis* YadA of the challenge bacteria are anticipated to show, by comparison with control immunized rabbits, a reduction in bacterial levels in plasma after challenge, with inhibition being still detectable in the plasma bacterial levels thereafter. This shows that bacterial infection was inhibited in the presence of antibodies to the YadA protein and suggests that a similar effect would prevail in other infected mammals.

Subjects infected with *Yersinia pestis* develop antibodies to YadA proteins and this is detected by ELISA and used to diagnose infection. Rabbits serums are tested prior to infection and 8 weeks after infection. All pre-infection serums were negative and all 8 week post infection serums were positive.

EXAMPLE 5

A. **Results: Amino acid sequence analysis of YadA and related proteins**

Sequence analysis of YadA, UspA1, and UspA2 showed that their stalks are most likely formed by extended coiled-coil domains (Fig. 1A). However, the coiled-coil forming probabilities for YadA were surprisingly low, prompting us to search for unusual features in this sequence. A Fast Fourier analysis of the putative coiled-coil segment (Fig. 1B) revealed a strong 15-residue periodiocity with the highest harmonic peak at 3.75 (15/4) resulting from a set of degenerate 15-residue repeats recognizable in the sequence (Fig. 2B). Secondary structure prediction and hydrophobic moment analysis suggested that the entire repeat region forms a strongly amphipathic α-helix, in agreement with the coiled-coil analysis. The observed periodicity of 3.75 residues per turn is significantly larger than the 3.5-3.6 typically observed in left-handed coiled coils (Seo and Cohen, 1993) or the 3.67 postulated for right-handed coiled coils (Peters *et al.*, 1996). Structurally, it is best compatible with a tightly supercoiled right-handed coiled coil having a pitch of 11.5 nm, a pitch angle of approximately 20°, and a length of about 17.5

nm (as compared to 17 nm measured in electron micrographs). In contrast, the main periodicity of the UspA1 and A2 stalk sequences is 3.52, suggesting a canonical left-handed coiled coil.

Fast Fourier analysis of the putative YadA head sequence also revealed a periodic structure, with a repeating pattern of approximately 14 residues (Fig. 1B). In the sequence, this was recognizable as a succession of degenerate repeats containing an alternating pattern of branched-chain aliphatic and small residues, followed by a position consisting mainly of Ala, Gly, Ser, or Thr (Fig. 2A). The same periodicity and repeat pattern was found in the head sequence of UspA1, but not of UspA2 (Fig. 2A). Secondary structure prediction suggested that this repeat region consists primarily of β-strands.

Sequence comparisons between YadA, UspA1, and UspA2 showed that all three sequences have a similar C-terminal domain (Fig. 2C). In addition YadA and UspA1 have similar head sequences and UspA1 and UspA2 have similar stalk sequences, giving UspA1 the appearance of a mosaic protein. Searches in GenBank and in the unfinished genomes database at NCBI yielded a surprising number of sequences that were clearly related to YadA and UspAs, from a phylogenetically diverse set of free-living and pathogenic proteobacteria (Table 1 and Table 2). Several of these sequences appear to be frameshifted (including the YadA homologue in *Yersinia pestis*), suggesting that they represent inactive genes. The similarity was most pronounced in a short sequence element, which is found in YadA between the head and stalk repeats and which we therefore named the 'neck' (Fig. 2C). About a quarter of its positions are practically invariant and another quarter highly conserved, making this - to our knowledge - the most highly conserved motif in any family of outer membrane proteins. Most genes contain a single copy of the neck sequence, but occasionally, up to 10 copies can be observed.

All YadA-like sequences have a conserved C-terminal region, which presumably anchors these proteins to the outer membrane (Fig. 2C). It consists of a short coiled-coil segment (Fig. 1A) and four transmembrane β-strands, as judged from secondary structure prediction and comparisons to a profile of porin β-strands (Baldermann *et al.*, 1998; A. Lupas and H. Engelhardt, unpublished). The β-strands are most similar to the equivalent C-terminal strands of eight-stranded porins (Baldermann *et al.*, 1998) and autotransporters (Loveless and Saier, 1997), which include many adhesins

such as *Yersinia* Ail, neisserial opacity proteins, *Escherichia coli* AidA, and *Bordetella* pertactin (Fig. 2C). The similarity includes a nearly invariant glycine, but the reasons for its conservation are unclear to us and do not appear to be illuminated by the recently published structure of the OmpA transmembrane domain (Pautsch and Schulz, 1998).

B. Discussion: Domain organization of YadA

The purification, immunolabeling and electron microscopic studies of *Y. enterocolitica* cells showed that the non-fimbrial, outer membrane adhesin YadA forms "lollipop"-shaped projections on the bacterial surface. Based on (i) the architecture of native and mutant YadA-oligomers, (ii) functional expression studies of N-terminal and C-terminal truncated YadA (Tamm *et al.*, 1993; Roggenkamp *et al.*, 1995 and 1996) and (iii) the amino acid sequence of the YadA monomers, three different domains in the molecules can be distinguished: A C-terminal outer membrane anchor domain, a rod-like intermediate segment and an N-terminal oval domain involved in adhesion to cells and ECMs, and auto-agglutination.

Tamm *et al.*, (1993) have demonstrated that the C-terminus of YadA carries a typical OMP-sorting signal and that C-terminally-truncated YadA is not located within the outer membrane. Our sequence analysis of the C-terminus predicts four amphiphatic transmembrane β-strands. As this part of the molecule is not visible in negatively stained cells or cell envelopes, it is likely to be completely buried in the outer membrane. The hydrophobic character of the anchor domain is illustrated by the tendency of the YadA oligomers to form small vesicles or even large, membrane-like layers via this domain. On the other hand, the amphipathic nature of the β-strands suggests that they form a solvent-accessible pore in the YadA oligomer. If so, one may envisage that - like autotransporters - YadA-like proteins mediate their own passage through the outer membrane. This possibility has several interesting implications. For example, a pore of sufficient size to translocate a polypeptide chain would require at least a trimer, probably a tetramer. Once the export step was completed, the polypeptide segments connecting the transmembrane β-strands to the conserved coiled-coil domain would be running through the pore, presumably in extended conformation, with the coiled-coil domain plugging the outer opening of the pore. It is interesting to note in this context that the C-terminal end of the

coiled-coil domain is hydrophobic and mainly composed of a residue with a minimal side-chain - alanine (Fig. 2C). The conserved nature of the entire C-terminal region would imply that all proteins of the YadA family have the same membrane-bound structure and therefore also the same number of subunits per oligomer.

Table 1 Genes homologous to YadA in public databases

	abbreviation	origin	source	no. of genes [a]	incom-plete	frame-shifted	remarks
• proteobacteria							
Rhizobacteria							
Sinorhizobium meliloti 1021	Sme	genomic	Stanford [c]	2	2		putative fragments of the same gene
• proteobacteria							
• thiobacilli							
Thiobacillus ferrooxidans ATCC23270	Tfe	genomic	TIGR [d]	2	2		putative fragments of the same gene; contains a unique repeat motif
Neisseriaceae							
Neisseria gonorrheae FA1090	Ngo	genomic	Oklahoma U [e]	2	1	1	incomplete gene appears disrupted by insertion of a multicopy genetic element
Neiseria meningitidis A Z2491[b]	Nma	genomic	Sanger [f]	1			head domain homologous to Haemophilus Hia/Hsf proteins
Neiseria meningitidis B MC58[b]	Nmb	GenBank		2			Hia/Hsf homolog and a second gene not present in Nma; neither present in Ngo
• proteobacteria							
Enterobacteriaceae							
Escherichia coli ECOR-9	Eco	GenBank		4			Eib genes are carried by Atlas prophages; phage P-EibA may be defective
Salmonella enteritidis LK5	Sen	genomic	U of Illinois [g]	1	1		same gene in all four Salmonella species; contains 10 copies of the neck seq.
Salmonella paratyphi A ATCC9150	Spa	genomic	Washington U [h]	1		1	gene in Sp contains 8 frameshifts
Salmonella typhi CT18	Sti	genomic	Sanger	1		1	
Salmonella typhimurium LT2	Sty	genomic	Washington U	1		1	
Yersinia enterocolitica	Ye	GenBank		2			from plasmids pYV6471/76 and pYV8081
Yersinia pestis CO-92 biovar Orientalis	Ypc	genomic	Sanger	5		2	YadA gene is frameshifted; two YadA-like genes contiguous on contig 653
Yersinia pseudotuberculosis	Yps	GenBank		1			from plasmid pIB1
Pasteurellaceae							
Actinobacillus actinomycetemcomitans	Aac	genomic	Oklahoma U	3	2		
Haemophilus ducreyi 35000	Hdu	GenBank		1			DsrA ; not present in H. influenzae Rd
Haemophilus influenzae nontypable 11	Hi1	GenBank		1			Hia
Haemophilus influenzae C54	Hi2	GenBank		1			Hsf (Hia homolog)

Haemophilus influenzae Rd [b]	Hi3	GenBank		2	1	1	Hia gene is frameshifted; incomplete gene resembles YadA
Pasteurella multocida PM70	Pmu	genomic	U of Minnesota [i]	2			Pmu 70b contains a unique repeat motif
Pseudomonas group							
Moraxellaceae							
Moraxella catarrhalis O35E	Mca	GenBank		2			UspA2 has a unique head sequence

[a] This column includes incomplete and frameshifted genes
[b] Complete genomes
[c] http://cmgm.stanford.edu/~mbarnett/genome.htm
[d] http://www.tigr.org/
[e] http://www.genome.ou.edu/
[f] http://www.sanger.ac.uk/Projects/Microbes/
[g] http://www.salmonella.org/
[h] http://genome.wustl.edu/gsc/Projects/bacteria.shtml
[i] http://www.cbc.umn.edu/ResearchProjects/AGAC/Pm/index.html

TABLE 2

organism	strain	contig/orf	copy	sequence
Sinorhizobium meliloti (Smc)	1021	423093G11	1	VTNVANGTVAKDSKDAVNGGQL
			2	ITNVAAGDLNANSTDAVNGSQL
			3	IANVAKG---VKATDAVNVGQL
		423017B11	4	ITGVAAG---TADSDAAHVAQL
Thiobacillus ferrooxidans(Tfe)	ATCC 23270	contig3426	1	IINVAPGTLSQTSTDAVNGSQL
			2	ITNVAPG---TTPNDVATVGQL
Neisseria gonorrheae(Ngo)	**FA1090**	**contig 18**		IVGVDDG---VNDFDAVNVRQL
		contig 20		IVNVADG---TEDLDAVNVRQL
Neisseria meningitidis[1](Nmc)	serogroup A strain Z2491	NMA1200		ITNVAPG---VKEGDVTNVAQL
	serogroup B strain MC58	NMB0992		ITNVAPG---VKEGDVTNVAQL
Escherichia coli(Eco)	ECOR-9	EibC		LGGLSDG---TRNSDAATVGQL
		EibD		LGGLSDG---TRNSDAATVGQL
		EibE		LGGLTDG---TRPSDAATVRQV
Salmonella typhimurium(Stm)	**LT2**	**contigs 1413, 553,1099,429**	**1**	ITNLAAGTLAADSTDAVNGSQL
			2	ITNLAAGTLAADSTDAVNGSQL
			3	ITNLAAGTLAADSTDAVNGSQL
			4	ITNVAAGELSEESTDAVNGSQL
			5	ITNVAAGALSEDSTDAVNGSQL
			6	ITNVAAGEIASDSTDAVNGSQL
			7	ITNVADG---SEAQDAVTVRQL
			8	ITNVAAS---SAEIDAVNVGQL
			9	ITNVAAG---VNNTDAVNVAQL
			10	IGNVSAA---VNDTDAVNYAQL
Salmonella typhi (Sty)	**CT18**	**contig 429**	**1**	ITNLAAGTLAADSTDAVNGSQL
			2	ITNLAAGTLAADSTDAVNGSQL
			3	ITNLAAGTLAADSTDAVNGSQL
			4	ITNVAAGELSEESTDAVNGSQL
			5	ITNVAAGALSEDSTDAVNGSQL
			6	ITNVAAGEIASDSTDAINGSQL
			7	ITNVADG---SEAQDAVTVRQL
			8	ITNVAAG---SADTDAVNVGQL
			9	ITNVAAG---VNNTDAVNVAQL
			10	IGNVSAA---VNDTDAVNYAQL
Salmonella paratyphi A (Spa)	**ATCC 9150**	**19646**	**1**	ITNLAAGTLAADSTDAVNGSQL
			2	LTNLAAGTLAADSTDAVNGSQL
			3	ITNLAAGTLTADSTDAVNGSQL
			4	ITNVAAGELSEESTDAVNGSQL

			5	ITNVAAGALSEDSTDAVNGSQL
			6	ITNVAAGEIASDSTDTVNGSQL
			7	ITNVADG---SEAQDAVTVRQL
			8	ITNVAAG---SADTDAVNVGQL
			9	ITNVAAG---VNNTDAVNVAQL
			10	IGNVSAA---VNDTDAVNYAQL
Salmonella enteritidis	**LK5**	**contig 606**	**6**	ITNVAAGEIASDSTDAVNGSQL
			7	ITNVADG---SEAQDAVTVRQL
			8	ITNVAAG---SADTDAVNVGQL
			9	ITNVAAR---VNNTDA>
Yersinia enterocolitica (Yc)	plasmid pYV6471/76	YadA		LTHLAAG---TKDTDAVNVAQL
	plasmid pYV8081	YadA		LTHLAAG---TKDTDAVNVAQL
Yersinia pseudotuberculosis(Yps)	plasmid pIB1	YadA		LTHLAAG---TEDTDAVNVAQL
Yersinia pestis(Ypc)	CO-92 Biovar Orientalis	**contig 3 (YadA)**		LTHLAAG---TEDTDAVNVAQL
		contig 653_1		ITGVKAG---VADTDAANVGQL
		contig 653_2		ITSVKKG---VENTDTINVSQL
		contig 721	**1**	ITGVAAG---SADYDVVNVAQL
			2	ITGVAAG---SADYDAVNVAQL
			3	VSNVADG---REAMDAVNLRQL
			4	LTNIAVG---TQGTDAVNLDQL
		contig 670	2	ITGVAAG---SASSDAVNVAQL
			3	VSNVADG---LQATDAVNLRQL
			4	ITNVAPA---TQGTDAVNFDQL
Actinobacillus actinomycetemcomitans(Aac)	HK1651	contig 220	1	ITQVADG---VNDKDAVNKSQL
			2	ITGVKDGEISAHSKEAVNGSQL
			3	IANVAKG---KAPTDAVNMSQL
		contig 439	1	IINVAGG---RNDTDAVNIAQL
			2	IKNVAAGNVAANSTDAVNGSQL
			3	ITGVAEG---TQDDDAVNFKQL
			4	ITNVADGVIAANSKDAVNGGQL
Haemophilus influenzae(Hin)	nontypeable strain 11	Hia		IDNVANGDISATSTDAINGSQL
	C54	Hsf		IDNVANGEISATSTDAINGSQL
	Rd KW20	**HI1732**		**VKNVAAGEISATSTDAINGSQL**
		HI1460		VANVKTA---EKDTDAVNKKQM
Pasteurella multocida (Pmu)	PM70	pm70b	1	LVNVADG---VDEQDAATVAQL
			2	VQNVASGLVTEASTDAINGSQL
			3	ITNVAAG---QNETDAVNVKQL
			4	ITNVTAG---VDDKDAVNVSQL
			5	VTGVADGDISPNSTDAVNGSQL
			6	IDGVADGEVNATSKQAVNGSQL
		pm70a	1	ITNVAAGRIDKTSTDAINGSQL
			2	VTNVQNGNIAKDSKDAVNGGQL

Moraxella catarrhalis [2](Mca)	O35E	UspA1	3	ISGIAKGDISENSTDAVNGSQL IVHVGAGEISDTSTDAVNGSQL

[1] identical neck sequence in Hsf/Hia orthologs from strains 1000, 205900, 297-0, 2996, 528, 860800, 90/18311, 93/4286, A22, BZ133, BZ147, BZ169, BZ198, BZ232, BZ83, E26, E32, F6124, H44/76, MC58, NG3/88, NG6/88, NGE28, NGE31, NGF26, NGH15, NGH36, NGH38, NGP165, SWZ107, Z2491

[2] identical neck sequence in UspA1 orthologs from strains TTA37,TTA24,O12E,O46E,V1171,ATCC25238

Bolded genes are frameshifted and unlikely to be coding

It will be clear that the invention may be practiced otherwise than as particularly described in the foregoing description and examples. Numerous modifications and variations of the present invention are possible in light of the above teachings and, therefore, are within the scope of the appended claims.

WHAT IS CLAIMED IS:

1. An isolated polypeptide conserved in proteobacterial extracellular domains comprising the sequence:

Arg- X^1 - X^2 -Thr- X^3 - X^4 -Ala- X^5 - Gly- X^6 - X^7 - X^8 - Thr-Asp-Ala-Val-Asn- X^9 - X^{10} -Gln-Leu [SEQ ID NO: 1],

wherein X^1 is selected from the group consisting of Gln, Lys, Thr, Val, and Arg;

wherein X^2 and X^4 are independently selected from the group consisting of Leu, Ile, and Val;

wherein X^3 is selected from the group consisting of His, Gly, Ser, Asn, and Gln;

wherein X^5 is selected from the group consisting of Ala, Lys, Val, Asp, Pro, Asn, Gly, and Glu;

wherein X^6 is selected from the group consisting of Thr, Val, Ser, Arg, Leu, Gln, Asp, Glu, Lys, and Asn;

wherein X^7 is selected from the group consisting of Lys, Glu, Ala, Gln, Ile, Asn, and Val;

wherein X^8 is selected from the group consisting of Asp, Asn, Gly, Ala, Ser, and Pro;

wherein X^9 is selected from the group consisting of Val, Leu, Phe, Gly, Lys, Met, and Ile; and

wherein X^{10} is selected from the group consisting of Ala, Gly, Ser, Asp, Arg, and Lys.

2. The sequence according to claim 1, wherein X^2 is Ile and X^4 is Val.

3. The sequence according to claim 1, which is selected from the group consisting of:

(a) Arg Gln Leu Thr His Leu Ala Ala Gly Thr Lys Asp Thr Asp

Ala Val Asn Val Ala Gln Leu [SEQ ID NO: 2];

(b) Arg Gln Leu Thr His Leu Ala Ala Gly Thr Glu Asp Thr Asp Ala Val Asn Val Ala Gln Leu [SEQ ID NO: 3];

(c) Arg Gln Leu Thr Asn Ile Ala Val Gly Thr Gln Gly Thr Asp Ala Val Asn Leu Asp Gln Leu [SEQ ID NO: 4]; and

(d) Arg Lys Ile Thr Gly Val Ala Ala Gly Ser Ala Asp Thr Asp Ala Val Asn Val Ala Gln Leu [SEQ ID NO: 10].

4. An isolated polypeptide sequence conserved in proteobacterial extracellular domains selected from the group consisting of sequences of the formula:

Arg Gln Ile Thr X^1 Val Lys X^2 Gly Val X^3 X^4 Thr Asp X^5 X^6 Asn Val X^7 Gln Leu [SEQ ID NO: 6],

wherein X^1 and X^7 are independently Gly or Ser; X^2 is Ala or Lys; X^3 is Ala or Glu; X^4 is Asp or Asn; X^5 is Ala or Thr; and X^6 is Ala or Ile.

5. An isolated polypeptide sequence conserved in proteobacterial extracellular domains selected from the group consisting of sequences of the formula:

Arg Lys Ile Thr Gly Val Ala Ala Gly Ser Ala X^1 X^2 Asp X^3 Val Asn Val Asn Gln Leu [SEQ ID NO: 8],

wherein X^1 is Asp or Ser; X^2 is Tyr or Ser; and X^3 is Val or Ala.

6. An isolated polypeptide sequence conserved in proteobacterial extracellular domains selected from the group consisting of sequences of the formula:

Arg Thr Val Ser Asn Val Ala Asp Gly X^1 X^2 Ala X^3 Asp Ala Val Asn Leu Arg Gln Leu [SEQ ID NO: 12], wherein X^1 is Arg or Leu; X^2 is Glu or Gln; and X^3 is Met or Thr.

7. An isolated polypeptide sequence conserved in proteobacterial extracellular domains selected from the group consisting of sequences of the formula:

Val Val Ile Asp Asn Val Ala Asn Gly X^1 Ile Ser Ala Thr Ser Thr Asp Ala Ile Asn Gly Ser Gln Leu [SEQ ID NO: 26], wherein X^1 is Asp or Glu.

8. An isolated polypeptide sequence conserved in proteobacterial extracellular domains of the formula selected from the group consisting of

(a) Lys Arg Ile Ala Asn Val Ala Lys Gly Lys Ala Pro Thr Asp Ala Val Asn Met Ser Gln Leu [SEQ ID NO: 17];

(b) Arg Arg Ile Ile Asn Val Ala Gly Gly Arg Asn Asp Thr Asp Ala Val Asn Ile Ala Gln Leu [SEQ ID NO: 18];

(c) Asn Arg Ile Thr Gly Val Ala Glu Gly Thr Gln Asp Asp Asp Ala Val Asn Phe Lys Gln Leu [SEQ ID NO: 19];

(d) Arg Gln Ile Lys Asn Val Ala Ala Gly Asn Val Ala Ala Asn Ser Thr Asp Ala Val Asn Gly Ser Gln Leu [SEQ ID NO: 20];

(e) Lys Lys Ile Thr Asn Val Ala Asp Gly Val Ile Ala Ala Asn Ser Lys Asp Ala Val Asn Gly Gly Gln Leu [SEQ ID NO: 21];

(f) Arg Lys Ile Val Gly Val Asp Asp Gly Val Asn Asp Phe Asp Ala Val Asn Val Arg Gln Leu [SEQ ID NO: 22];

(g) Arg Gln Ile Thr Asn Val Ala Pro Ala Thr Gln Gly Thr Asp Ala Val Asn Phe Asp Gln Leu [SEQ ID NO: 23];

(h) Arg Gln Ile Val Asn Val Gly Ala Gly Gln Ile Ser Asp Thr Ser Thr Asp Ala Val Asn Gly Ser Gln Leu [SEQ ID NO: 24]; and

(i) Gly Arg Ile Thr Gln Val Ala Asp Gly Val Asn Asp Lys Asp Ala Val Asn Lys Ser Gln Leu [SEQ ID NO: 25].

9. The composition according to claim 1 wherein said polypeptide is produced synthetically.

10. The composition according to claim 1 wherein said polypeptide is produced recombinantly.

11. A fusion protein comprising a polypeptide of any of claims 1-10 fused in frame to a second protein.

12. The fusion protein according to claim 11 wherein said second protein is selected from the group consisting of an *E. coli* DnaK protein, a GST protein, a mycobacterial heat shock protein 70, a diphtheria toxoid, a tetanus toxoid, galactokinase, ubiquitin, α-mating factor, β-galactosidase, and influenza NS-1 protein.

13. An isolated polynucleotide sequence encoding a polynucleotide of any of claims 1-10.

14. A nucleic acid molecule comprising a polynucleotide sequence of claim 13, said polynucleotide sequence under the control of regulatory sequences which direct the expression of said polypeptide in a host cell.

15. A recombinant virus comprising a polynucleotide sequence of claim 13, said polynucleotide sequence under the control of regulatory sequences which direct the expression of said polypeptide in a host cell infected by said virus.

16. A host cell comprising the nucleic acid molecule of claim 14 or the virus of claim 15.

17. A composition which inhibits or retards the binding of the polypeptide of any of claims 1 to 10 to its ligand or to a cell expressing its ligand.

18. An antibody which binds to the polypeptide of any of claims 1- 10.

19. The antibody according to claim 18 which is selected from the group consisting of a polyclonal antibody, a monoclonal antibody, a chimeric antibody, a recombinant antibody, a humanized antibody, a human antibody, a Fab fragment thereof, a Fab_2 fragment thereof, an F_v fragment thereof; and mixtures thereof.

20. An antibody which is an anti-idiotype of the antibody of claim 18.

21. An immunogenic composition useful as a vaccine to prevent infection by a proteobacterial species comprising in a pharmaceutically acceptable carrier, at least one component selected from the group consisting of:

(a) a polypeptide of any of claims 1-10 or an immunogenic fragment thereof;

(b) an amino acid sequence at least 70% identical to the sequence of (a) as determined by a sequence comparison algorithm, which sequence binds the ligand of (a);

(c) a small molecule which binds the ligand of (a);

(d) an antibody which binds (a);

(e) an anti-idiotype antibody of (d); and

(f) a fusion protein comprising a polypeptide of (a) fused in frame to a second protein;

and an optional adjuvant.

22. A process for producing a polypeptide comprising culturing the host cell of Claim 16 under conditions suitable for expression of said polynucleotide sequence, and isolating from the cell or cell lysate a polypeptide encoded by said polynucleotide sequence.

23. A method for vaccinating a patient against infection with a proteobacteria comprising administering to the patient a prophylactically effective amount of the composition of claim 21.

24. A method of making an immunogenic composition for use as a vaccine component against proteobacterial infection comprising:

fusing a polypeptide of any of claims 1-10 to a second protein capable of resisting degradation *in vivo*, wherein said polypeptide elicits antibodies *in vivo* which interfere with the binding of the bacterial adhesin molecules to their receptors.

25. A process for diagnosing a bacterial infection comprising contacting

a biological sample from a possibly infected subject with a labeled antibody which binds to the conserved polypeptide of any of claims 1 to 10; and measuring the signal generated by said label with a suitable assay, wherein detection of said signal indicates the presence of an adhesin molecule from said bacteria.

26. A diagnostic reagent which comprises a composition capable of binding to a polypeptide of any of claims 1-10, said composition associated with a detectable label.

27. A method for identifying compounds which antagonize the binding of a bacterial adhesin to its ligand comprising the steps of:

providing a sample of said ligand or a cell which expresses said ligand immobilized on a support;

contacting said sample with a known amount of a polypeptide of any of claims 1 - 10 and a known amount of a test compound;

washing unbound materials from said sample;

contacting said sample with a labeled reagent which binds to said polypeptide;

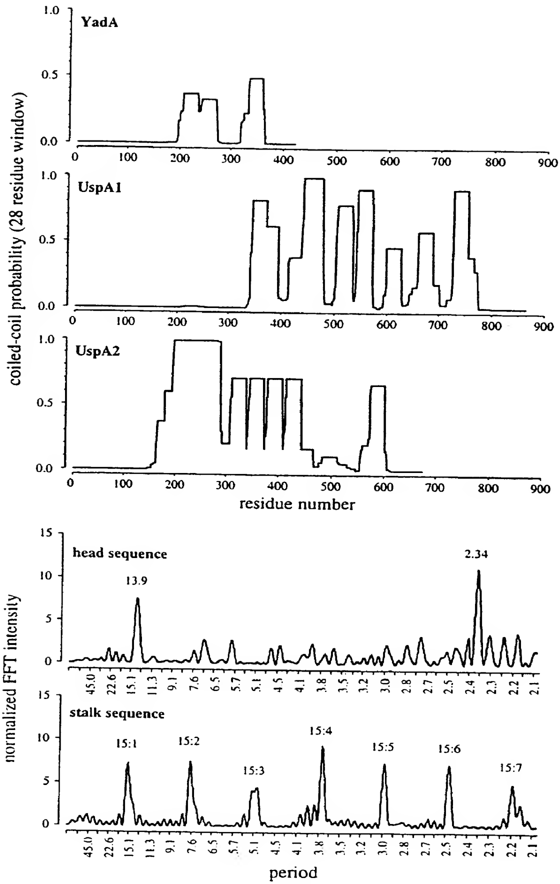
washing unbound reagent from said sample;

measuring the amount of signal generated by said label, wherein the amount of signal generated is inversely proportional to the ability of the test compound to disrupt or inhibit binding between said polypeptide and said ligand; and

identifying those test compounds as antagonists which are associated with a low signal.

28. A method for generating a small molecule which antagonizes the binding between a proteobacterial adhesin and its ligand comprising analyzing an antibody to a polypeptide of any of claims 1-10 in a computer modelling program.

FIGURE 1



SUBSTITUTE SHEET (RULE 26)

```
A<<<------------------------------------------------------------------------- head ------------------------------------
                                     ------------------------><-- neck >>>
Sme         ---------------------------------------------------------------------------------------------------------
ITVGADPADPNAPTTVITGGS--IVIGSTMLGSNGL---VITGGPSVTTDGIDAGGMKVTNVANG
Tfe         ---------------------------------------------------------------------------------------------------------
TAVGVNATSTGTNS---------VALGYNSSDGGQNNV-VAVGSATQQR--------KIINVAPG                      LGLGSAVTQYGDSNN-IASGPGSAAGTND---
Ngo 18      IALGKQAKATKQNG-MALGVESEAAGDFS--TAVGNESKAKGQGG--VGLGNQSKAEADFA--------VAVGNKAEATKENS--LVIGRYARANGNHS--
VSLGSRSEIKDGVSNS-------VAPGYGSVASENNV--VSVAYKETPQSTELSYR-KIVGVDDG
Ngo 20      -------------------------------VALGFRSSHNLFA--SSFGAFSTATAIES--------LAVGDSSQSTGYRS--ATFGSHSRALAEES--
LALGYETRANAYGS---------VALGAESVANEENT--VSVSSDTLKR--------KIVNVADG
Sty         ------------------------------------------------TAVGYDAVASGAGC--------LALGQNSSSSIEGS--IALGSGSTSNRA----
ITTGIRETSATSDG---------VVIGYNTTDRELLGA-LSLGTDGESYR-------QITNVADG
Yel YadA    73  IAIGATAEAAKGAA-VAVGAGSIATGVNS--VAIGPLSKALGDSA--VTYGAASTAQKDG---------VAIGARASTSDTG---VAVGFNSKADAKNS--
VAIGHSSHVAANHGYS-------IAIGDRSKTDRENS--VSIGHESLNR--------QLTHLAAG
Ye2 YadA    70  IAVGASAEAAKQAA-VAVGAGSIATGVNS--VAIGPLSKALGDSA--VTYGAASTAQKDG---------VAIGARAFTSDTG---VAVGFNSKVDAKNS--
VAIGHSSHVAVDHDYS-------IAIGDRSKTDRKNS--VSIGHESLNR--------QLTHLAAG
Yps YadA   108  IAIGATAEAAKPAA-VAVGSGSIATGVNS--VAIGPLSKALGDSA--VTYGASSTAQKDG---------VAIGARASASDTG---VAVGFNSKVDAQNS--
VAIGHSSHVAADHGYS-------IAIGDHSKTDRENS--VSIGHESLNR--------QLTHLAAG
Ype YadA        IAIGATAEAAKPAA-IAVGSGSMATGVDS--VAIGPLSKALGDSA--VTYGVSSTAQKDG---------VAIGAKASASDTG---VAVGFNSKVDAQNS--
VAIGHSSHVAADHGYS-------IAIGDHSKTDRENS--VSIGHESLNR--------QLTHLAAG
Ype 653a    ------------------------------VNIGDFNTGVIGNG--TAVGNNNSLGGSTNG-------VVVGNGGSLSNSING-VVIGNGSVSDGDG---
VSVGGGTSTNGG-----------IAIGSGSNATRSDE--MNIGDR------------QITGVKAG
Ype 653b    ---------------------------------------------------------------------VSVSGNGTSGNNV---VNIGGAANGNNSAS--
LGTGSVSSEGG------------IALGSGSIATRNDE--LNIGDR------------QITSVKKG
Ype 721     ---------------IAVGPATVVNGNNG--IGIGSSAVVGPSAVGGIAIGPNTQATGIAS--------TALGAGSQAHGSQS--LALGAGATASQANS--
IALGASSVTTVGAESDY------SAYGLTAPQTSVGE--VGMGTAQGNR--------KITGVAAG
Ype 670     ---------------IAVGPATVVSGDNG--VGIGNTALVGAAATGGIAIGFGTQVTAAGA--------TAIGSAAQAQGAQS--LALGAGAVTSQANS--
IALGAASINTVGAQSSY------SAYALTAPQASVGE--LGIGTALGNR--------KITGVAAG
Aac 439     ---------------IAVGTSALAQKESA--IAFGYQANASGINA--ISLGANAKASQDNV--------VAIGKYATATESGS--MAIGQGAKSTFKNS--
LALGTGTIVNSVDGGQS------KFTAQNYDANNGV---VAVANAGKER--------RIINVAGG
Pmu 70a     ---------------IAFGSEAKTTENAGSSIAFGTKAQTRASAS--IAIGMGAETGFDGGQALDGSDAVALGREAKAKRQNA--LAFGYKAVADHKDA--
VALGAGAETAAAEGTNEATVNEFKYSGFAGIKPIAT---VSVGKKDAER--------TITNVAAG
Pmu 70b     ------------------------------TVVGAQTKAASRS---TALGAEATALGLDA--------VAVGITSKAEGDKS--VAIGANSTADSNN---
VALGATSRAVAKEGSG-------YLTGKQSSLV------VSVGHQDGAADQHILR--RLVNVADG
Mca UspA1 177  TVAGGYNNEATGTDSTIAGGRKNQATGKGS-FAAGIDNKANADNA--VALGNKNTIEGENS--------
VAIGSNNTVKKGQQNVFILGSNTDTTNAQNGSVLLGHNTAGKAATIVNSAEVGGLSLTGFAGASKTGNGT-VSVGKKGKER--------QIVHVGAG
```

FIGURE 2A
2/4

```
B
                <------------------------------------------------------------ stalk ------------------
--------------------------------------------->
                                   repeat 1 |     repeat 2 |      repeat 3 |     repeat 4 |     repeat 5 |     repeat 6 |
   repeat 7 |      repeat 8 |          repeat 9 |
Sme            ------------------------------------------------------------------------------------------
ESVAESKSYTDEKTEWAIDQAAIYTDQVIETKVSAVNNYAQQR------------------
Yel YadA  219
EKTQENTNKRSAELLANANAYADNKSSSVLGIANNYTDSKSAETLENARKEAFAQSKDVLNMAKAHSNSVARTTLETAEEHANSVARTTLETAEEHANKKSAEALASANVYAD
SKSSHTLKTANSYTDVTVSNSTKKAIRESNQYTDHK  367
Ye2 YadA  216  EKTQVNANKKSAE--------------VLGIANNYTDSKSAETLENARKEAFDLSNDALDMAKKHSN--------------
SVARTTLETAEEHPNKKSAETLARANVYADSKSSHTLQTANSYTDVTVSNSTKKAIRESNQYTDHK  334
Yps YadA  254  ----------------------------------AETLENARKETLAQSNDVLDAAKKHSN--------------
SVARTTLETAEEHANKKSAEALVSAKVYADSNSSHTLKTANSYTDVTVSSSTKKAISESNQYTDHK  346
Ype YadA       ----------------------------------AETLENARKETLAQSNDVLDAAKKHSN--------------
SVARTTLETAEEHANKKSAETLVSAKVYADSNSSQTLKTANSYTDVTVSNSTKKATRESNQYTDHK
Ype 653a       ----------
GETLNSANIYVDNQATETLNNANIYTDNKATETINNANTYTDNKSSETLNSANSYTDNKSSETLNSANTYTDSKTAEIFNTTKTYMDGKSKETLNNTYDYVDSKVSSIVYDVN
SYTDKTVNTAFETSLSDAKSYVDDK
Ype 653b       ----------
DDVLNLSNEYSDNSFSTVTENINNYTDASLDTVLNTTGEYTDNSILLVTNESNNYTDNGMESVSNYANIYADESLLAIYNEEANYMSNLIDVTLNNANNYTDLSVNTIIYTGK
QYTDSRINEYQRTFKNEFLTYSNGK

C
               <---- neck region ---->               <--- coiled coil segment --->          <-------------------
membrane anchor ------------------>    <-coiled coil segment->
                                                     hhhhhhhhhhhhhhhhhhhhhhhhhhhhhh              sssssss   ssssssss
sssssssss            sssssssssss
Sme            LIANVAKG---VKATDAVNVGQLD.......[stalk]...FAQLSGEIGQVRSEARQAAAIGLAAASLRFDNEPG—KLSVALGGGFWRSEGALAFGAGYTS
-EDG>
Tfe            QITNVAPG---TTPNDVATVGQLQGA---------------VGQAQHYAAQVGSVNAAALNAAASAASGQGPN-------
TVAGGYGEYDGQSAFAFTYQHRFNCNWQALLTVGSNGSGKN-------TEVGAGASYSW
Ngo 18         KIVGVDDG---VNDFDAVNVRQLKAM---------------QGQNMAELFSVRSEVRGVAASSAALSALTPLSYDANNPTQFMVGFGTYKGRQAMALGLSHF--
VNDRFMVKVGGTLGGNKTG-----HMMNVGLTWKFG[16]LMQDNRQLKPRLEKLEARLNAM
Ngo 20         KIVNVADG---TEDLDAVNVRQLNR------------------------LSKRTNRVGASAAALASLKPAQLGKNDKFAFSLGFGSYKNAQAVAMGAVFK--
PAENVLLNVAGSFAGPD-------RAFGAGVFWKFG[13]HPAKVLQLRQEVAALRARQAET
Nme            RITNVAPG---VKEGDVTNVAQLKGV---------------AQNLNNRIDNVDGNARAGIAQAIATAGLVQAYLPG--KSMMAIGGGTYRGEAGYAIGYSSIS-
DGGNWIIKGTASGNSRGH------FGASASVGYQW
Stm            RIGNVSAA---VNDTDAVNYAQLKRSVEEANTYTDQK----MGEMNSKIKGVENKMSGGIASAMAMAGLPQAYAPG--ANMTSIAGGTFNGESAVAIGVSMVS-
ESGGWVYKLQGTSNSQGD------YSAAIGAGFQW
Sty            RIGNVSAA---VNDTDAVNYAQLKRSVEEANTYTDQK----MGEMNSKIKGIENKMSGGIASAMAMAGLPQAYAPG--ANMTSIAGGTFNGESAVAIGVSMVS-
ESGGWVYKLQGTSNSQGD------YSAAIGAGFQW
Spa            RIGNVSAA---VNDTDAVNYAQLKRSVEEANTYTDQK----MGEMNSKIKGVENKMSGGIASAMAMAGLPQAYAPG--ANMTSIAGGTFNGESAVAISVSMVS-
ESGGWVYKLQGTSNSQGD------YSAAIGAGFQW
Yel YadA  195  QLTHLAAG---TKDTDAVNVAQLKKEI....[stalk]...FRQLDNRLDKLDTRVDKGLASSAALNSLFQPYGVG--KVNFTAGVGGYRSSQALAIGSGYR--
VNENVALKAGVAYAGSSD------VMYNASFNIEW   455          gi|401465
```

FIGURE 2B
3/4

```
Ye2 YadA  192  QLTHLAAG---TKDTDAVNVAQLKKEI....[stalk]...FRQLDNRLDKLDTRVDKGLASSAALNSLFQPYGVG--KVNFTAGVGGYRSSQALAIGSGYR--
VNESVALKAGVAYAGSSD------VMYNASFNIEW   422            gi|96988
Yps YadA  230  QLTHLAAG---TEDTDAVNVAQLKKEM....[stalk]...FSQLDNRLDKLDKRVDKGLASSAALNSLFQPYGVG--KVNFTAGVGGYRSSQALAIGSGYR--
VNESVALKAGVAYAGSSN------VMYNASFNIEW   434            gi|141104
Ype YadA       QLTHLAAG---TEDTDAVNVAQLKKEM....[stalk]...FSQLDNRLDKLDKRVDKGLASSAALNSLFQPYGVG--KVNFTAGVGGYRSSQALAIGSGYR--
VNESVAFKAGVAYAGSSN------VMYNASFNIEW
Ype 653a       QITGVKAG---VADTDAANVGQLVAKA....[stalk]...YNQLSDKVNKNFNKTNAGISGAMAMSGIPQKFGY---EKSFGMAIGAYRGQSALAVGGDWN--
INHKTITRVNVSADTEGG------VGVAAGFAFGIN
Ype 653b       QITSVKKG---VENTDTINVSQLNDSF....[stalk]...FGGFDKDINQKQKQLNAGIAATMAAAVIPQKSGS---KVSIGVGLAGYSDQGAGSVGAIWH--
VNQRITMNTTMTYDTQRG------VSLLTGLSIGI
Ype 721        QLTNIAVG---TQGTDAVNLDQLNHSMSNVTNDANAYTDQRYSALKEDLKKQDSTLSAGIAGAMAMASLTQPYTPG--ASMATIGAASYRGQSALSVGVSSIS-
DSGRWVSKLQASSNTQGD------MGVGVGVGYQW
Ype 670        QITNVAPA---TQGTDAVNFDQLKSISNQTNAYTNQR----YSELKQDLRKQNSVLSAGIASAMSMASLTQPYTSG--SSMTTIGAASYRGQSALSLGVSSIS-
DSGRWVSKLQASSNTQGD------FGIGVGVGYQW
Aac 220        RIANVAKG---KAPTDAVNMSQLQDVGSAINNR-------IDNIDKRVKKMDKRRKAGTASALATAGLMQPHRDG--QSALVAAVGQYQSETAVAVGYSRIS-
DNGKYGVKVSFSTNSQGE------VGGTAGAGYFW
Aac 447                                      <[stalk]...IDRLDSRVNELDKEVKNGLASQAALSGLFQPYNVG--SLNLSAAVGGYKSKTALAVGSGYR--
FNQNVAAKAGVAVSTNGGS-----ATYNVGLNFEW
Hin Hia    974  VIDNVANGDISATSTDAINGSQLYAVAKGVTNLAGQ-----VNNLEGKVNKVGKRADAGTASALAASQLPQATMPG--KSMVAIAGSSYQGQNGLAIGVSRIS-
DNGKVIIRLSGTTNSQGK------TGVAAGVGYQW  1098            gi|1235666
Hin Hsf   2229  VIDNVANGEISATSTDAINGSQLYAVAKGVTNLAGQ-----VNNLEGKVNKVGKRADAGTASALAASQLPQATMPG--KSMVAIAGSSYQGQNGLAIGVSRIS-
DNGKVIIRLSGTTNSQGK------TGVAAGVGYQW  2353            gi|1666683
Hin 1732       AVKNVAAGEISATSTDAINGSQLYAVAKG-----------VTNLAGQVNKVGKRADAGTASALAASQLPQASMPG--KSMVSIAGSSYQGQSGLAIGVSRIS-
DNGKVIIRLSGTTNSQGK------TGVAAGVGYQW                 gi|1073822
Pmu 70a        KISGIAKGDISENSTDAVNGSQLYELQQKIAKSGDN-----YNILNNRINKVDKDLRAGIAGANAAAGLPQAYIPG--KSMVAVAAGTYKGQNAIALGMSRIS-
DNGKVIIKLTGNTNSRGD------FGASIGAGYQW
Pmu 70b        VIDGVADGEVNATSKQAVNGSQLHKVQQQVNNQATA-----INKLGDHINKVDKDLRAGIAGATAVAFLQRPNEAG--KSIVSLGVGSYRSESAIAVGYARNS-
DNNKISIKLGGGMNSRGD------VNFGGSIGYQW
Mca UspA1 315  QIVHVGAGEISDTSTDAVNGSQLHALA....[stalk]...GQHFNNRISAVERQTAGGIANAIAIATLPSPSRAG--EHHVLFGSGYHNGQAAVSLGAAGLS-
DTGKSTYKIGLSWSDAGG------LSGGVGGSYRWK  863            gi|5453174
Mca UspA2      -------(missing)-----------....[stalk]...VNAFDGRITALDSKVENGMAAQAALSGLFQPYSVG--KFNATAALGGYGSKSAVAIGAGYR--
VNPNLAFKAGAAINTSGNK----KGSYNIGVNYEF   674            gi|5453184

SSSSSSSSS           SSSSSSSSSS                                                                                    SSSSSSSSS
Eco OmpA                                                                                                          VSPVFAGGVEYA--
ITPEIATRLEYQWTNNITRP[8]NGMLSLGVSYRF                 gi|129135
Yen Ail                                                                                                           TSMAYGAGVQFN--
PLPNFVIDASYEYSKLDSIKV---GTWMLGAGYRF                 gi|113536
Ngo OpaA                                                                                                          LGFGAVAGVGID--
ITPKLTLDAGYRYHNWGTRF[5]THEASLGVRYRF                 gi|417406
Eco AidA                                                                                                          NQGEIKTGIEGV--
ITQNLSVNGGVAYQAGGHGSN--AISGALGIKYSF                 gi|543788
Bpe P.68                                                                                                          TRAELGLGMAAA--
LGRGHSLYASYEYSKGPKLAM--PWTFHAGYRYSW                 gi|464364
```

FIGURE 2B Cont.

4/4

SEQUENCE LISTING

<110> SMITHKLINE BEECHAM CORPORATION

<120> CONSERVED ADHESIN MOTIF AND METHODS OF USE THEREOF

<130> GM50047

<140> TO BE ASSIGNED
<141> 2000-04-13

<150> US 60/129,073
<151> 1999-04-13

<160> 26

<170> FastSEQ for Windows Version 3.0

<210> 1
<211> 21
<212> PRT
<213> Artificial Sequence

<220>
<221> unsure
<222> (2)(3)(5)(6)(8)(10)(11)(12)(18)(19)
<223> Derived from Yersinia specias

<400> 1

```
Arg Xaa Xaa Thr Xaa Xaa Ala Xaa Gly Xaa Xaa Xaa Thr Asp Ala Val
 1               5                  10                  15
Asn Xaa Xaa Gln Leu
            20
```

<210> 2
<211> 21
<212> PRT

<213> Yersinia enterocolitica

<400> 2

Arg Gln Leu Thr His Leu Ala Ala Gly Thr Lys Asp Thr Asp Ala Val
1 5 10 15
Asn Val Ala Gln Leu
20

<210> 3
<211> 21
<212> PRT
<213> Yersinia pseudotuberculosis

<400> 3

Arg Gln Leu Thr His Leu Ala Ala Gly Thr Glu Asp Thr Asp Ala Val
1 5 10 15
Asn Val Ala Gln Leu
20

<210> 4
<211> 21
<212> PRT
<213> Yersinia pestis

<400> 4

Arg Gln Leu Thr Asn Ile Ala Val Gly Thr Gln Gly Thr Asp Ala Val
1 5 10 15
Asn Leu Asp Gln Leu
20

<210> 5
<211> 21
<212> PRT
<213> Artificial Sequence

<220>
<221> unsure
<222> (5)(8)(11)(12)(15)(16)(19)
<223> Dervied from Yersinia species

<400> 5

Arg Gln Ile Thr Xaa Val Lys Xaa Gly Val Xaa Xaa Thr Asp Xaa Xaa
1 5 10 15
Asn Val Xaa Gln Leu
20

<210> 6
<211> 21
<212> PRT
<213> Yersinia pestis

<400> 6

Arg Gln Ile Thr Gly Val Lys Ala Gly Val Ala Asp Thr Asp Ala Ala
1 5 10 15
Asn Val Gly Gln Leu
20

<210> 7
<211> 21
<212> PRT
<213> Yersinia pestis

<400> 7

Arg Gln Ile Thr Gly Val Lys Lys Gly Val Glu Asn Thr Asp Thr Ile
1 5 10 15
Asn Val Ser Gln Leu
20

<210> 8
<211> 21
<212> PRT
<213> Yersinia pestis
<220>
<221> unsure
<222> (12)(13)(15)

<400> 8

Arg Lys Ile Thr Gly Val Ala Ala Gly Ser Ala Xaa Xaa Asp Xaa Val

1 5 10 15
Asn Val Asn Gln Leu
20

<210> 9
<211> 21
<212> PRT
<213> Yersinia pestis

<400> 9
Arg Lys Ile Thr Gly Val Ala Ala Gly Ser Ala Asp Tyr Asp Val Val
1 5 10 15
Asn Val Asn Gln Leu
20

<210> 10
<211> 21
<212> PRT
<213> Yersinia pestis

<400> 10
Arg Lys Ile Thr Gly Val Ala Ala Gly Ser Ala Asp Tyr Asp Ala Val
1 5 10 15
Asn Val Asn Gln Leu
20

<210> 11
<211> 21
<212> PRT
<213> Yersinia pestis

<400> 11
Arg Lys Ile Thr Gly Val Ala Ala Gly Ser Ala Ser Ser Asp Ala Val
1 5 10 15
Asn Val Asn Gln Leu
20

<210> 12
<211> 21

<212> PRT
<213> Yersinia pestis
<220>
<221> unsure
<222> (10)(11)(13)

<400> 12
Arg Thr Val Ser Asn Val Ala Asp Gly Xaa Xaa Ala Xaa Asp Ala Val
1 5 10 15
Asn Leu Arg Gln Leu
20

<210> 13
<211> 21
<212> PRT
<213> Yersinia pestis

<400> 13
Arg Thr Val Ser Asn Val Ala Asp Gly Arg Glu Ala Met Asp Ala Val
1 5 10 15
Asn Leu Arg Gln Leu
20

<210> 14
<211> 21
<212> PRT
<213> Yersinia pestis

<400> 14
Arg Thr Val Ser Asn Val Ala Asp Gly Leu Gln Ala Thr Asp Ala Val
1 5 10 15
Asn Leu Arg Gln Leu
20

<210> 15
<211> 24
<212> PRT
<213> Haemophilus influenzae

<400> 15

Val Val Ile Asp Asn Val Ala Asn Gly Asp Ile Ser Ala Thr Ser Thr
1 5 10 15
Asp Ala Ile Asn Gly Ser Gln Leu
20

<210> 16
<211> 24
<212> PRT
<213> Haemophilus influenzae

<400> 16

Val Val Ile Asp Asn Val Ala Asn Gly Glu Ile Ser Ala Thr Ser Thr
1 5 10 15
Asp Ala Ile Asn Gly Ser Gln Leu
20

<210> 17
<211> 21
<212> PRT
<213> Actinobacillus actinomycetemcomitans

<400> 17

Lys Arg Ile Ala Asn Val Ala Lys Gly Lys Ala Pro Thr Asp Ala Val
1 5 10 15
Asn Met Ser Gln Leu
20

<210> 18
<211> 21
<212> PRT
<213> Actinobacillus actinomycetemcomitans

<400> 18

Arg Arg Ile Ile Asn Val Ala Gly Gly Arg Asn Asp Thr Asp Ala Val
1 5 10 15
Asn Ile Ala Gln Leu
20

<210> 19
<211> 21
<212> PRT
<213> Actinobacillus actinomycetemcomitans

<400> 19

Asn Arg Ile Thr Gly Val Ala Glu Gly Thr Gln Asp Asp Asp Ala Val
1 5 10 15
Asn Phe Lys Gln Leu
20

<210> 20
<211> 24
<212> PRT
<213> Actinobacillus actinomycetemcomitans

<400> 20

Arg Gln Ile Lys Asn Val Ala Ala Gly Asn Val Ala Ala Asn Ser Thr
1 5 10 15
Asp Ala Val Asn Gly Ser Gln Leu
20

<210> 21
<211> 24
<212> PRT
<213> Actinobacillus actinomycetemcomitans

<400> 21

Lys Lys Ile Thr Asn Val Ala Asp Gly Val Ile Ala Ala Asn Ser Lys
1 5 10 15
Asp Ala Val Asn Gly Gly Gln Leu
20

<210> 22
<211> 21
<212> PRT
<213> Neisseria gonorrheae

<400> 22

Arg Lys Ile Val Gly Val Asp Asp Gly Val Asn Asp Phe Asp Ala Val
1 5 10 15

Asn Val Arg Gln Leu
20

<210> 23
<211> 21
<212> PRT
<213> Yersinia pestis

<400> 23

Arg Gln Ile Thr Asn Val Ala Pro Ala Thr Gln Gly Thr Asp Ala Val
1 5 10 15

Asn Phe Asp Gln Leu
20

<210> 24
<211> 24
<212> PRT
<213> Moraxella catarrhalis

<400> 24

Arg Gln Ile Val Asn Val Gly Ala Gly Gln Ile Ser Asp Thr Ser Thr
1 5 10 15

Asp Ala Val Asn Gly Ser Gln Leu
20

<210> 25
<211> 21
<212> PRT
<213> Actinobacillus actinomycetemcomitans

<400> 25

Gly Arg Ile Thr Gln Val Ala Asp Gly Val Asn Asp Lys Asp Ala Val
1 5 10 15

Asn Lys Ser Gln Leu
20

<210> 26

<211> 24
<212> PRT
<213> Artificial Sequence

<220>
<221> unsure
<222> (10)
<223> Derived from Yersinia species

<400> 26
Val Val Ile Asp Asn Val Ala Asn Gly Xaa Ile Ser Ala Thr Ser Thr
1 5 10 15
Asp Ala Ile Asn Gly Ser Gln Leu
20

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/09866

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) :Please See Extra Sheet.
US CL :Please See Extra Sheet.
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 424/130.1, 184.1; 435/7.1, 69.3, 235.1, 325; 530/300, 350, 387.8, 388.1; 536/23.7

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EAST, MEDLINE, BIOSIS, CA, CAPLUS, EMBASE
terms: polypeptide, conserved, bacterial, adhesins

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	SKURNIK et al. Analysis of the yopA gene encoding the Yop1 virulence determinants of Yersinia spp. Molecular Microbiology. 1989, Vol. 3, pages 517-529, see entire document	1-3
X ----- A	WO 98/28333 A2 (THE BOARD OF REGENTS, THE UNIVERSITY OF TEXAS SYSTEM) 02 July 1998 (02-07-98), see entire document, especially claim 22.	8-28 ----- 4-7

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
16 AUGUST 2000	07 SEP 2000

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231
Facsimile No. (703) 305-3230

Authorized officer
MARK NAVARRO
Telephone No. (703) 308-0254

A. CLASSIFICATION OF SUBJECT MATTER:
IPC (7):

A61K 38/00, 39/00, 39/395; C07H 21/04; C07K 1/00, 16/00; C12N 5/00, 7/00, 15/09; C12P 21/08; G01N 33/53

A. CLASSIFICATION OF SUBJECT MATTER:
US CL :

424/130.1, 184.1; 435/7.1, 69.3, 235.1, 325; 530/300, 350, 387.8, 388.1; 536/23.7